E ERIS Project, Final Report

Name of Call Area Bidding For (tick ONE only):	
☐ **Strand A1: Automated metadata generation and text mining** ☐ **Strand A2: Developing e-infrastructure to support research disciplines** ☐ **Strand A3: Repositories: start-up** ☐ **Strand A4: Repositories: rapid innovation** ☑ **Strand A5: Repositories: enhancement** ☐ **Strand A6: Preservation exemplars** ☐ **Strand B1: VRE Innovation: Tools and interoperability** ☐ **Strand B2: VRE Innovation: VRE Frameworks** ☐ **Strand B3: VRE Innovation: VRE National and Institutional Interoperability**	
Name of Lead Institution:	University of Edinburgh on behalf of SCURL (Scottish Confederation of University and Research Libraries)
Name of Project:	ERIS: **E**nhancing **R**epository **I**nfrastructure in **S**cotland
Name(s) of Project Partner(s): Primary Partners: University of Edinburgh (lead institution); CDLR (Centre for Digital Library Research, University of Strathclyde); DCC (Digital Curation Centre); National Library of Scotland; OCLC / RLG; SAC (Scottish Agricultural College); SCURL (Scottish Confederation of University and Research Libraries); SLIC (Scottish Library and Information Council); University of Glasgow; **Secondary Partners**: ***Research pools:*** SAGES (Scottish Alliance for Geoscience, Environment and Society), SICSA (Scottish Informatics and Computer Science Alliance); SIRE (Scottish Institute for Research in Economics); SULSA (Scottish Universities Life Sciences Alliance) ***SCURL members****:* Heriot-Watt University; RGU (The Robert Gordon University); UWS (University of the West of Scotland); University of Aberdeen; University of Dundee; University of St Andrews; University of Stirling; The project also enjoys the full support of Universities Scotland.	
Full Contact Details for Primary Contact: Name: Sheila Cannell Position: Director of Library Services, University of Edinburgh Email: sheila.cannell@ed.ac.uk Tel: 0131 650 3383; Fax: 0131 650 3380 Address: 5 Buccleuch Place, University of Edinburgh, Edinburgh, EH8 9LJ	
Length of Project:	24 months
Project Start Date:	1 April 2009
Total Funding Requested from JISC:	

Funding Broken Down over Financial Years (April - March):	
April 09 – March 10	**April 10 – March 11**
£164,966	£182,869

Total Institutional Contributions:

Outline Project Description: The conceived purpose of the ERIS project was to develop – in close partnership with researchers and their institutions' repository managers – a set of user-led and user-centric solutions that will motivate researchers to deposit their work in repositories, facilitate the integration of repositories in research and institutional processes and, as a result, develop the original IRIScotland pilot into a trusted cross-repository resource discovery service, capable of providing access to a critical mass of Scottish research output. In order to achieve this overall aim, ERIS paid particular attention to the requirements of research pooling – an innovative cross-institutional way of conducting research, which has been widely credited for having substantially contributed to Scotland's RAE 2008 successes.

Table of Contents

ERIS PROJECT, FINAL REPORT 1

TABLE OF CONTENTS 2
ACKNOWLEDGEMENTS 4
EXECUTIVE SUMMARY 5
WORKPACKAGES 6
BACKGROUND 7
AIMS AND OBJECTIVES 8
METHODOLOGY 8
IMPLEMENTATION 9
OUTPUTS AND RESULTS 10
SCOS 10
RESEARCH POOL ACTIVITY 10
OUTCOMES 11
CONCLUSIONS. 11
IMPLICATIONS 12
RECOMMENDATIONS 12
POST ERIS ACTIVITIES 13

APPENDICES 16

APPENDIX A: DIGEST OF KEY POINTS FROM WP1 ACTIVITY 17
INSTITUTIONAL PERSPECTIVE/RESEARCH POOL PERSPECTIVES 17
RESEARCHER PERSPECTIVE 17
CONCLUSION 18
SCOS 19
RESEARCH POOL ACTIVITY 19
APPENDIX B: WP1 - ERIS FOCUS GROUPS: REPORT 20
APPENDIX C: NEW MODES OF RESEARCH DISSEMINATION AND SUPPORT 23
KEY FINDINGS 26
CONCLUSION 31
APPENDIX D: ERIS PROJECT NLS WORK-PACKAGE 3 ACTIVITY 46
PURPOSE 47
MISSION 47
BACKGROUND 47
AIMS AND OBJECTIVES 47
MEASURES OF SUCCESS 49
TESTING 50
ASSUMPTIONS, DEPENDENCIES AND CONSTRAINTS 50
ASSUMPTIONS 50
DEPENDENCIES 50
CONSTRAINTS 50
RESOURCE CONSTRAINTS 50
TIME CONSTRAINTS 50
NON- FINANCIAL RESOURCE CONSTRAINTS 50
LEGAL/POLICY CONSTRAINTS 51
PROJECT INTERFACES 51
PROJECT GOVERNANCE 51
ROLES AND RESPONSIBILITIES: 51
COMMUNICATIONS 52
PROJECT CONTROLS 52

COSTS 52
COSTS AND RESOURCES 52
SCHEDULE 52
APPENDIX E: WORK PACKAGE 3 - TECHNOLOGICAL ENHANCEMENTS 52
APPENDIX F: UWS/SAC BASIC BRIEF 59
APPENDIX G - ERIS PROJECT - NLS WORK-PACKAGE 3 ACTIVITY/CDLR DEVELOPMENT SCOPE (EPRINTS USER VIEWS) 61
PURPOSE 62
MISSION 62
BACKGROUND 62
AIMS AND OBJECTIVES 62
MEASURES OF SUCCESS 63
DELIVERY APPROACH 63
TESTING 65
ASSUMPTIONS, DEPENDENCIES AND CONSTRAINTS 65
ASSUMPTIONS 65
DEPENDENCIES 66
CONSTRAINTS 66
RESOURCE CONSTRAINTS 66
TIME CONSTRAINTS 66
NON- FINANCIAL RESOURCE CONSTRAINTS 66
SCOPE EXCLUSIONS 66
INTERFACES 66
GOVERNANCE 67
ROLES AND RESPONSIBILITIES: 67
COMMUNICATIONS 67
PROJECT CONTROLS 67
COSTS 67
COSTS AND RESOURCES 67
SCHEDULE 67
RISKS 68
APPENDIX H: ENHANCING REPOSITORY INFRASTRUCTURE IN SCOTLAND -WORK PACKAGE 3 – DEVELOPMENT REPORT (UNIVERSITY OF STRATHCLYDE) 69
PHASE 1 – USE CASES & SCENARIOS 69
PHASE 2 – DEVELOPMENT 69
SUBJECT METADATA 69
KEYWORDS 70
EMAIL REMINDERS 70
MEPRINTS WIDGETS 70
SOFTWARE USED 71
ISSUES & LESSON LEARNED 71
REFERENCES 71
APPENDIX I: ERIS WORK PACKAGE 2: CASE STUDIES 72
CONTENTS 72
PURPOSE AND INTENDED OUTCOMES 72
1. CURATION POLICY AT THE UNIVERSITY OF ABERDEEN 72
2. CURATION POLICY AT THE UNIVERSITY OF GLASGOW 75
3. CURATION POLICY AT THE UNIVERSITY OF ST ANDREWS 77
4. SUMMATION 80
APPENDIX J: ERIS QUALITY PLAN (AUTUMN 2010). WP 4. 82

Acknowledgements

The JISC programme for the ERIS project (Enhancing Repository Infrastructure in Scotland) was the Information Environment, and the strand under which it was funded was Repository enhancement. The primary partners for ERIS were the University of Edinburgh (lead institution); the CDLR (Centre for Digital Library Research, University of Strathclyde); the DCC (Digital Curation Centre); the National Library of Scotland; OCLC / RLG; SAC (Scottish Agricultural College); SCURL (Scottish Confederation of University and Research Libraries); and SLIC (Scottish Library and Information Council); the University of Glasgow.

Executive Summary

Unlike its predecessor project IRIScotland, which took a more top-down approach to developing repository services for users in Scottish HE institutions, the ERIS project took a bottom-up approach, and consulted potential users of repository systems widely at the outset of the project, whether these systems would be for use by single institutions or subject pools. The idea was to find out what the user community actually required in terms of repository systems and services.

In practice we discovered that there was a range of opinion about the usefulness of repositories to researchers, and to institutions. Some groups within HE are already clear about the usefulness of the technology, and the associated workflows. These do not need much persuading. Others see the suggestion that they should use repositories to disseminate their research as an imposition of extra work, additional to the existing publication process, either for themselves or their research administrators. This spectrum of opinion surfaced in both the focus group activity and the case studies, reports on which form part of the appendices to this report. Perhaps surprisingly, the division of opinion was also found among the research pools and their users.

The staff of the project responded to the spectrum of opinion, and the range of commitment to the role of repositories in dissemination, and in the research process, by reformulating their ideas about engagement with the research community in Scotland, and about what was required to achieve the goal of enhancing repository infrastructure for users and stakeholders. This was done in an agile way, and the new approach and new targets were formulated and agreed, and reflected in the quality plan which can be found at Appendix J.

The new approach consisted of two strands. The first was about increased advocacy for the use of repositories, which involved the offer of hosted repository services during the later stages of the project, for institutions which did not already have their own institutional repositories. If these (generally) smaller institutions wished to continue to have access to these hosted repository spaces after the end of the project, it was proposed that these repository-lite services would be administered by the SDLC (the Scottish Digital Library Consortium), at minimal cost. The idea of this was to ensure at least a minimal level of provision nationally, to encourage the use of repositories.

The second strand involved looking beyond the limited period of the project, to what would be required to promote the take-up and development of repository services in the longer term. It was concluded that this would involve developing the business case for continuing ERIS/IRIScotland type work. What the project had in mind was setting up co-ordinating groups and fora for meetings of stakeholders and users, and the research pools. This would be the responsibility of the SCOS group (the Scottish Council for Open Scholarship). It was planned that SCOS would be launched formally at a event closure meeting, held in the Royal Society of Edinburgh, in September 2011. [Quality Plan Appendix J].

The course of the project after the community consultations followed this new approach. The project continued to be agile in response, and influenced related work in Scotland (see section: 'Outcomes').

We have a number of recommendations to make for future approaches to repository operations in Scotland. These are listed in a section following.

Workpackages

Wp1 - Focus group activity. This was conducted by Susan Ashworth (U of G), and John MacColl reported on survey work; James Toon reported on research Pool activity. These reports are available in the appendices to this report. [Appendices A – C]

Wp2: Case studies. These are from Glasgow University and the DCC, on practical considerations around ERIS policy framework documents. [Appendix I-L]

Wp3a –Several strands to the Meprints software development (Strathclyde), where open source plug-ins have been developed for eprints users, allowing users to be emailed when papers are deposited, and regular mailing of viewing and download statistics. [Appendices G -H]

Wp3b – Research Pools aggregation. Early discussion with researchers and the managers of the Scottish research pools revealed that there was, in the short term at least, likely to be little call for aggregations of metadata across national institutions. There are too few subject records in Scottish repositories for such an aggregation to be worthwhile for the pools, and in any case, a national base would not be a first port of call for current research. [Appendix D]

Wp3c - Aggregation of records for the management of research by universities however would be useful however, and work towards a national aggregation was undertaken. There is potential here to use aggregated repository data to help researchers make contacts in emerging cross-disciplinary fields. If more research is exposed as metadata, then perhaps it would encourage sharing, and encourage more full text deposit as a result. There may be an issue about where the aggregation occurs, and the longer term relationship between the CRIS (which has emerged in a big way during the ERIS lifespan) and the IR. In terms of research management, we think that there would be huge value in aggregation to support joint REF submissions centred on pools.

This remains as work in progress, due to issues around the ways in which repositories report their contents. With a growing number of universities implemented CRIS systems with a common CERIF metadata model, this should be possible from these systems rather than the Institutional Repositories.

Wp3d – The ERIS project has effectively underwritten a national repository architecture which might underpin such an aggregation, were it to be desirable. ERIS has provided some smaller Scottish HE institutions, SAC and UWS, with a repository-lite implementation of DSpace on a trial basis up until the close of the project. After this trial period, it will be possible for these institutions to run the

repository-lite implementation indefinitely with low-cost support from the SDLC. Feedback from SAC and UWS is being collected, and will be available from the project wiki at the end of the project. [Appendix E]

Wp4 became the principal focus of the project after the initial consultation work, and it was, and is, about developing the business case for continuing ERIS/IRIScotland work, the setting up of co-ordinating groups and fora for meetings of stakeholders and users (SCOS group, Research Pool meetings). SCOS will be launched formally at a event closure meeting in September 2011. [Quality Plan Appendix J]

Background

ERIS – Enhancing Repository Infrastructure in Scotland - came into being as a successor project to the two phases of the IRIScotland project (Institutional Repository Infrastructure for Scotland), which ran between October 2005 and March 2008. The purpose of IRIScotland was to provide an aggregation service for institutional repositories, and a search service for that aggregation, so that it would be possible to cross-search the open access research output of all Scottish IRs. This demonstration was successful, and led to discussion of a project which would build on IRIScotland's outputs.

The two IRIScotland projects suggested that top-down advocacy, guidelines and the development of standards, while essential, are not sufficient by themselves to engage researchers with repositories and therefore to create the critical mass that repositories need in order to be seen as an essential part of the infrastructure by the researchers.

HE institutions around the world are beginning to mandate their academics to deposit their research outputs in Open Access institutional repositories. In Scotland the movement is rapidly gathering momentum partly as a result of the Scottish Declaration on Open Access and the work of IRIScotland, but it is still the case that only a minority of Scottish HEIs has so far introduced mandatory open access policies. To translate these policies into reality, we realised that Scottish HEIs needed to develop researcher-friendly repositories that fit neatly into the research workflow. The success of cross-repository services will dependent on this.

The ERIS project proposers took the view that to achieve high deposit levels in repositories it would be necessary to work in close collaboration with both researchers and research managers to reach a better understanding of how repositories fit into the research workflow, and how the addition of extra repository functionality may contribute to generating significantly higher deposit rates.

Repository services must be matched to the needs of researchers to ensure they offer real incentives to deposit*[1]. We know that, for example, researchers would value a service helping them to keep their personal bibliographies up to date; this could be set up by the repositories in such a way that the updating would only occur in conjunction with the deposit of a new research output. Other added-value services

[1] Association of Research Libraries (ARL). *The Research Library's Role in Digital Repository Services*. January 2009. p. 8. See at http://www.arl.org/bm~doc/repository-services-report.pdf.

may include tools for group work and version control at file level – particularly useful for research pools –, facilities for the deposit of other types of information including research datasets or for improving the visibility of research outputs amongst the business community. Essentially researchers would appreciate a workspace which exists between them and the repository service, so that the repository becomes a part of the research workflow. These are only examples.

Aims and Objectives

The purpose of the ERIS project was to develop – in close partnership with researchers and their institutions' repository managers – a set of user-led and user-centric solutions intended to motivate researchers to deposit their work in repositories, and to facilitate the integration of repositories in research and institutional processes and, as a result, develop the IRIScotland pilot into a trusted cross-repository resource discovery service, capable of providing access to a critical mass of Scottish research output.

Research pooling was a particular focus of the ERIS project. Research Pooling is an innovative cross-institutional way of conducting research, which has been widely credited with having substantially contributed to Scotland's RAE 2008 successes.

The focus of the project changed after the initial consultation with stakeholders and focus groups (wp1). The technical development seemed of less importance, given that the stakeholders and the researchers had expressed uncertainty about why they would want to have certain facilities, including aggregated metadata for research pool activity. The focus therefore shifted to some extent to the provision of advocacy, training, and the development of a business model for post ERIS activity, which would include an organised structure to promote and build on open access development in Scotland, not just for research pools, but also for repositories in general.

The principle of open access continues to be a strong driving force in Scotland, as well as the UK. The context in which it is of use is changing however, and it is seen as something which needs to be more firmly embedded in the actual research process. The ERIS engagement with the user community has highlighted that this is so, especially now that many institutions are engaging with research management tools, such as Atira's PURE. It is hoped that the ERIS outputs will be useful for institutions and individuals who wish to build on the insights they contain, and that SCOS may help to steer institutions through a fast changing repository and research management landscape.

Methodology

The project has been involved in survey and Focus group work to establish real attitudes to the use of open access repositories in Scotland, and to establish attitudes to using repositories as part of the research management process. Other survey work, completed by the end of the project (extended to July 2011) includes a survey of SCURL members to find out where they see repositories being placed in the future.

Case Studies and Use cases

A number of case studies from across Scotland were conducted in the course of the project, with the help of the Digital Curation Centre. Reports on these are available in the appendices. Use cases were put together by the project partners, exploring likely and desirable scenarios for repository deposit by researchers. These also will be available from the project wiki at the close of the project. [Appendix I]

SCOS

The project is in the process of setting up, as part of its sustainability planning, the *Scottish Council for Open Scholarship*. This body, consisting initially of those on the ERIS project board, will continue to move the open access agenda forward in Scotland after the project has closed. There will be an archived mailing list to serve this group, once it is formally inaugurated, at: openscholar@jiscmail.ac.uk.

Implementation

The original proposal was converted into a project plan, and the technical development was specified broadly in line with PRINCE 2 methodology. Successful outcomes were specified, and there was a risk assessment for each development.

The survey and focus group activity was conducted first (WP1), and the responses had a significant impact on the rest of the project. Some of the objectives of the project did not resonate fully with the requirements of the target group (principally researchers and research pools). Case studies and Policy documents were the focus of WP2, and these were written throughout the project, and completed in early 2011. Technical development was affected (as mentioned elsewhere) by the responses to the survey and the focus groups of WP1, and the work on the aggregation of metadata was delayed until late in the project (autumn of 2010).

Shortly after the project began, the The WP3 manager at the NLS became the ERIS project manager, and no replacement was made. Hence long-term sustainability has now been seen as more appropriately channelled through SDLC than the NLS. This is partly due to legal deposit priorities changing, making NLS less focused on relevant TDR development, and partly as a result of the common sense around moving development into Edinburgh where it could be managed more effectively.

Workpackage 4 is essentially about creating a business plan for building on the activity of both the ERIS project, and its predecessor IRIScotland, and creating a series of recommendations for future work. A formal group is being created (Scottish Council for Open Scholarship) which will act as a driver for future Open Access and related research management issues.

Outputs and Results

After the survey work which was the core of Workpackage 1, it was concluded that it was important to:

1. To embed repositories in the institutional infrastructure - focus on provision of support to the research management process – collaboration with research offices especially important. Libraries take a leading role in services that ensure the record of research is maintained and recorded etc. (include the management of open access demands, such as OA funds, subject repositories) – Includes development of suitable and interoperable infrastructures (i.e. CERIF etc)

2. Focus on the development of open scholarship as a long term endeavour for researchers – starting at early career stage. 10 year plan….(hard/soft activity based around development of infrastructure and in advocacy/community etc).

3. Develop information literacy support services for research in respect of open scholarship. Probably related to developments in academic liaison. (RLG-SRD)

Notably – researchers don't seem to care at all for repository features (with the exception of public facing lists). These features should be for the repository management/research management

SCOS

- Project working to set up the SCOS group, aiming to place open scholarship in context in and across Scotlands research landscape, both institutional and SME , governmental etc.

- Council will establish a programme of activity that will support the long term development needs of institutions in open scholarship/open knowledge and lobby non-HE to improve the links and engagement between researchers and users

- SCOS will also seek to update the OATS declaration to focus on open scholarship, having taken stock of the activity/success/failures of the last

RESEARCH POOL ACTIVITY

We worked with the research pools – together with other institutions and working with EuroCRIS and PURE, implemented CRISPool project – providing data to support the research process and to help facilitate strategic research management.

We also discussed areas of common activity between pools as part of the PAN network, only tentatively at present, but with the aim of starting to consider if it is possible for pools to share administrative costs, knowledge transfer activities etc. The project identified areas of opportunity to develop with pools and via SFC.

The project is in the process of setting up some 'DSpace Lite' repositories for organisations such as the Scottish Agricultural College (SAC), and the University of the West of Scotland (UWS). These repository instances provide open access opportunities for smaller HE institutions, which might otherwise not be able to participate in open access dissemination. Other institutions in Scotland have either set up their own repositories, or have access to institutional repository space via the services provided by the SDLC (Scottish Digital Library Consortium). The SDLC is a ready-made vehicle for sustainability, and offers the prospect of then building on this through SCURL, and achieving for repositories what SHEDL has done for licenced content.

The provision of DSpace Lite services as part of ERIS project activity means that all institutions in Scotland who want an open access repository can have one. This expands the range of sources of the metadata which can aggregated by a search service.

Software development

Software development has been undertaken with Strathclyde University to produce enhancements to repository reporting services for the ePrints platform. The software plugins (available as open source software) allows users to be notified when new deposits occur, and to receive regular updates of viewing and download stats, available by email. This facility has now been implemented with Glasgow's Enlighten repository.

Outcomes

The project's principle target was to respond to the views of users and stakeholders about what was required in order to enhance repository infrastructure in Scotland. As noted already, the views of the users and stakeholders covered a wide spectrum, and there was a lack of consensus except within specific groups. Discussion of the usefulness of repository infrastructure and participation in these discussions was facilitated by the provision of hosted repository space.

Conclusions.

There is still significant scope for advocacy (in a changing environment), regularly updated training, as well as system development in consultation with stakeholders.

Implications

It is clear that researchers are not at the moment sure what they can gain from repository services. The OA agenda is still developing, and becoming increasingly complex. There are still big questions to address about gold vs green and full text vs metadata in terms of costs, publisher strategies, REF and CRIS development, provided either by HE institutions or by research pools. Repositories have a big role in future for the management of research information, and several Scottish universities are now implementing research management tools (PURE in particular), and these can be used alongside standard repository software to provide university administrations and research administrators with a much wider array of information. This sort of development is a major step forward for business intelligence in the HE environment.

Recommendations

The project has revealed a need for greater clarity between open access and research management agendas, and has exposed confusion between open access, and research management, research visibility and research impact, leading to a recommendation that work with the academic community is needed to clarify this.

- Our perception is that there is support for OA amongst researchers, but the policy environment in which they work is volatile
- There should be a recommendation to the Scottish Govt/research funders and universities (research policy community): this could include something about the national mandate in Denmark, focusing on research funders, aimed at covering the gaps where funding is not accompanied by a green OA mandate.
- There is a need to raise the OA discussion to the government level from the institutional level, where only so much can be achieved.
- We should seek to promote the impact of Scottish research, and to improve its dissemination beyond HE..
- After initial consultations with the user community, it became clear that many user groups were unclear as to why they might want to use repository services. The focus of the project therefore shifted to some extent to the provision of advocacy, training, and the development of a business model for post ERIS activity. This includes an organised structure to promote and build on open access development in Scotland, not just for research pools, but also for repositories in general.
- That a ‘Scottish Council for Open Scholarship’ would be a good outcome from the project, to co-ordinate discussion of important issues, and to facilitate meetings of researchers and stakeholders from across Scotland. This should build on the OATS declaration from 2004.
- We need to work to ensure all institutions either run a repository of their own, or have access to repository services provided by a third party, and to explore how this is best managed.
- A cost benefit analysis may be useful to establish the most effective model of repository usage for institutions, research pools, and other users.
- We need to continue to work on the development of fair scholarly publishing models and the publishing debate (gold vs green etc.).

- Further work needs to be done to establish whether or not the Scottish research community would benefit from some kind of central aggregation of metadata, available through one or more portals.
- We need to explore whether, in the long term, we need individual repositories for institutions and subject groupings, or a single Scottish repository, utilising cloud-based, shared services.

Post ERIS activities

At the end of the ERIS project, the Project Board should consider which ERIS and related activities we should carry forward into a post-ERIS environment, and who should own these activities. The Board is invited to discuss these issues—these discussions will form part of the ERIS outcomes, and act as an agenda for the future.

The overarching purpose of ERIS has been about examining the business case for a collaborative approach to the management and promotion of Scottish research. As a Board, we should think about how we can continue with this agenda. The ideas below are all about ensuring that Scottish research is visible in Scotland and globally on behalf of Scottish HEI and funders.

There will be a major opportunity to promote this agenda at the post-ERIS meeting in September. This meeting is provisionally entitled “Promoting Scotland’s research and researchers: new directions”. This seminar will have a number of high profile speakers. We will briefly consider the ERIS project, but focus much more on the future, in particular the launch of the Scottish Council of Open Scholarship, and the possibility of seeking funding to develop a project on promoting the impact of research.

Aggregator service

The project has not created an overall national infrastructure or aggregated “big portal” of Scottish research. It is possible to harvest for different purposes, using the same OAI-PMH protocol. The ERIS project did not push to create an aggregator service because there was no appreciable demand for such a discovery service. As the Board, we should confirm what we consider about such a portal. It seems more likely that a portal supporting disciplines is needed, particularly when allied to other outputs and methods of measuring impact.

We should consider for which purposes and for what audiences such an infrastructure would be useful, e.g. for finding research, for research visibility, for research management, to measure the impact of research success, for business intelligence or to showcase Scottish research.

A portal could be created of:

- Pan-Scottish research (HEIs and non HEIs)
- Pan-Scottish research (HEIs only)
- For pools or other collaborations

- To showcase the best of Scottish research (although this approach, unless fully automated, has particularly issues in how to choose the “cream of research”.

Open Scholarship agenda

In the post-ERIS environment, we should also be addressing the wider Open Scholarship agenda, in particular the launch of the Scottish Council on Open Scholarship which will aim to promote Open Research in Scotland.

Impact of research

We could seek funding on promoting the impact of research. One model may be based on scaling up to a Scottish level the work being carried out the ESPRC funded project on sustainable urban development, the ISSUES project http://www.urbansustainabilityexchange.org.uk/ISSUESPeople.htm
This project showcases research for policy makers and practitioners, and provides links between researchers and practitioners in order to enhance knowledge transfer and exchange. An appropriate action in this post-ERIS environment may be carry out an audit of knowledge dissemination in Scotland, perhaps looking particularly at the research pools.

Business models

The ERIS Board should ensure that there is a continued locus for creating business models and reflecting on costs for repositories, institutional repositories, joint repositories and scholarly communications. For example, there can certainly be a discussion about the value of independent local repositories in each institution. The question may be about the value of local institutional repositories in supporting both Open Access and research assessment—are there better ways of handling each of these needs? The ERIS Board may, for example, wish to produce a case for funding to further reflect on business models (from either SFC or JISC).

Management of Institutional repositories

There will be a continued need for a coordinating process for Institutional repositories in Scotland. This might include a number of provisions, e.g.:

- A grouping for Scottish Institutional Repository Managers, to provide training and networking opportunities
- Supporting liaison and subject librarians in their activities with subject communities in getting the right information to the right subject networks.
- A locus for decision-making on common data model which will allow for working together, using for example the JISC EXRI model http://exri.bris.ac.uk
- A locus for considering a cloud-based solution to repositories.
- Ensuring that all institutions (in HE and beyond) which want to have research repositories can (currently arranged through the SDLC repository-lite service, but there may be other models).

In the longer term, it may be that a different model of collaborative repository management may emerge within Scotland.

Stakeholders

There are a large number of stakeholders with an interest in this territory—researchers, research funders, research managers, policy makers, government. ERIS has come out of the library domain, and there is continued work in this domain. However, if the ERIS agenda is itself to make a real impact, this work now needs to embrace a much wider community.

Appendices

Appendix A: Digest of key points from WP1 activity

Note that this information was gathered from meeting with research pools, researcher interviews, researcher survey work and a number of focus group sessions.

Institutional perspective/research pool perspectives

- Institutional repositories, where considered successful and embedded, are generally successful when in the context of research information management
- For managers of research, the REF requirements are the clear driver, and accepted as such
- Impact management emerging as area of concern that isn't adequately supported or understood as of yet.
- The driver is the acquisition of metadata describing the research outputs of the institution, not in management of full text, open access works.
- Researchers do feel loyalty to institution, but primarily at the departmental or subject level. (i.e. research pools/groups/centres etc) suggesting this is the area to target for advocacy.
- Researchers are happy(ish) to contribute materials to repositories if they clearly provide value to the research management process for their area (i.e. to ensure sustainability for their pool/centre etc)
- Pools require improvements in infrastructure to support the process of research management and knowledge exchange etc. (i.e interoperability work, such as use of schemes like CERIF)

Researcher perspective

- The pressures that researchers are under to publish are significant and success relies heavily on the existing publishing model for scholarly communications.
- Researchers are keen to establish ways of recording impact in light of new REF requirements.
- Generally supportive of open access *in principle* but this is very different depending on the career stage of the researcher
- Open access is accepted as part of the requirements researchers have to meet these days, but this is normally in the context of funder or other compliance requirements
- Subject repositories are considered much more valuable and relevant to researchers, examples such as Repec/Arxiv/SSRN/Pubmed
- Little evidence exists which has convinced our sample researchers of any scholarly benefits of institutional repositories.
- There are some vocal sceptics in the community who are unlikely to be swayed to the value of IR (and are in clear opposition) and this is usually attributed to unnecessary systems, lack of usability and lack of time)

- Researchers say they would participate in open access schemes if asked to do so, but are realistic about compliance problems (although this is a little contradictory)
- Some concerns re the drive open access as a threat to sustainability for small, niche societies that may be funded by journal subscriptions (good genuine point – not all publishers are profit hungry corporations and libraries should respect these issues)
- 50% Researchers would be happy to receive information which provides greater evidence of the value of their work through citations and other bibliometric indicators. It is not clear what these should be however. Suspect researchers just want more bibliometric support.
- Researchers acknowledged value in all cases for accessibility to grey literature. Have libraries been guilty of focusing on singular deliverables and not the sum of research produced? Focus on the project not the output. (Note: sits well with CERIF core entities)
- Management of grey literature is a significant theme in all areas. There is an acknowledgement of the lack of skills and knowledge and general understanding of the issues surrounding how to curate and assess grey literature and other materials produced during the course of research. Opportunities here for specialist training and infrastructure development. Developments in this area would attract the attention of academic staff.
- In this respect, researchers are very risk averse
- Retrospective curation of research (for open scholarship) very unlikely – more possible for new projects, but the challenges are likely to be in developing the infrastructure and encouraging methods/practices at early stages of research career.
- Researchers do have genuine reasons for not making grey literature and other materials available, and accommodating these issues will be critical in order to gain support. (although some just don't want to – citing 'no time' as being the most common reason)
- Researchers were broadly negative about library research services and saw them as user unfriendly and of little value.
- Researchers demand more subject specific services and are likely to approach colleagues and other researchers in this respect.
- Researchers would be interested to support systems that engage with the management of their research interests – such as recommender systems, dynamic info about usage/metrics, web CV's and online profile pages etc. (Mendeley?)

CONCLUSION

Suggests overall requirement for three goal strategy

4. To embed repositories in the institutional infrastructure - focus on provision of support to the research management process – collaboration with research offices especially important. Libraries take a leading role in services that ensure the record of research is maintained and recorded etc. (include the management of open access demands, such as OA funds, subject repositories) – Includes development of suitable and interoperable infrastructures (i.e. CERIF etc)
5. Focus on the development of **open scholarship** as a long term endeavour for researchers – starting at early career stage. 10 year plan….(hard/soft activity based around development of infrastructure and in advocacy/community etc)

6. Develop information literacy support services for research in respect of open scholarship. Probably related to developments in academic liaison. (RLG-SRD)

Notably – researchers don't seem to care at all for repository features (with the exception of public facing lists). These features should be for the repository management/research management

---

SCOS

- Project working to set up the SCOS group, aiming to place open scholarship in context in and across Scotlands research landscape, both institutional and SME , governmental etc.
- Council will establish a programme of activity that will support the long term development needs of institutions in open scholarship/open knowledge and lobby non-HE to improve the links and engagement between researchers and users
- SCOS will also seek to update the OATS declaration to focus on open scholarship, having taken stock of the activity/success/failures of the last

RESEARCH POOL ACTIVITY

- Working with research pools – together with other institutions and working with EuroCRIS and PURE, implemented CRISPool project – providing data to support the research process and to help facilitate strategic research management. Currently working on a proposal as part of the current JISC 15/10 call for a CRISPool 2
- Also discussing areas of common activity between pools as part of the PAN network, only tentatively at present, but with the aim of starting to consider if it is possible for pools to share administrative costs, knowledge transfer activities etc…. project identified areas of opportunity to develop with pools and via SFC.

Appendix B: WP1 - ERIS Focus Groups: Report

The ERIS project hosted three focus groups for researchers to discuss their experiences and opinions of institutional repositories and related activity.

The focus groups were held at the Scottish Agricultural College on 22nd January 2010 (5 participants), and in Edinburgh on the 19th (3 participants) and 20th May (5 participants).

Scottish Agricultural College 22nd January 2010

Repositories
For researchers at SAC any publications arising from research funded by the Government or DEFRA should be made available in the public domain. Researchers are not always making publications and research data openly available, but they will provide it on demand. Many researchers put PDFs of journal articles onto personal websites. When asked what they think about making their publications openly available, one participant stated that we do the research, hand over the IP when we publish and then buy back the research, so why don't we tell publishers to take a hike and distribute the PDFs as we see fit?
The participants had mixed feeling about institutional repositories and were not immediately convinced of the benefits. Although they would probably deposit into an IR if mandated by their institution, there was a feeling that such compliance would be patchy. Some subject areas, such as economics, are more used to making ongoing research available and will of often have working papers series available on a website and have a culture more accepting of the citation of unfinished ideas and unpublished work.
One researcher stressed how useful it is as a user of research to be able to openly access publications. He is writing a textbook and has spent a huge amount of time trying to get hold of papers - repositories of openly available papers would have made this process much easier for him. It would not matter to him what he found was an author final version as long as he could cite the authoritative version. Another researcher said that she thought researchers would want the authoritative versions and as they can find the metadata easily enough they can simply email an author and ask for a copy. So what use is the repository?

Open access
One participant was particularly concerned about the impact of open access on his learned society. The income from journal subscriptions funds the activity of the society. There are three journals which currently have various embargo periods but the editorial board are under pressure to move to six month embargoes for all of the journals and he was concerned that, if this happens, libraries would cease subscriptions.
Other participants felt that anything that increases the chances of a paper getting cited is a good thing but as not all downloads lead to a citation it might be useful if researchers could use downloads as an additional metric. It also was felt that access by members of the public to science – 'citizen science' is very important and should be emphasised.
Some of the participants were particularly interested in the use of repositories for hosting grey literature. Researchers at SAC generate a huge amount of applied, non peer-reviewed research materials, it would be very valuable to have a repository for this kind of material and it would help researchers to record and demonstrate the impact of this research.

19th May 2010

Participants

i) Researcher in Psychology and manager of a research centre. Has a requirement to report annually to the MRC about the research outputs of members of the centre (papers, lectures, conference proceedings). Members of the research centre are not informing their manager about what they have published, never mind depositing publications into an institutional repository.
ii) Contract researcher and PhD student. Interested in access to information and acquiring knowledge in a very time pressured environment. Finds the research process frustrating.

It is a struggle to get funding, a struggle to engage research participants and a struggle to publish. A new experience is the cost to publish (asked to pay Biomed Central publication charges of £995). Where does money come from, particularly if grants are not coming from the larger RCUK/Wellcome routes.

iii) Senior Lecturer in Anaesthetics. Interested in access to information – especially PhD/MSc/grey literature.

Pressure to publish

The participants were asked about the pressures on them to publish. There is huge pressure to apply for grants and prestigious publications help grant applications; pressure in universities is to bring in grant funding; pressure from Research Council to disseminate anything that's publicly funded. One participant pointed out that while it is important to get published it is also important to disseminate findings and there is also work that is important that journals do not necessarily want to publish and where can researchers place this work.

Academic researchers aim to publish in journals with high impact factor and there is a set of journals in each subject that they always aim for first. Level of output is also important so, for example, basic scientists in psychology compare themselves with social scientists although the outputs for basic scientists take longer to achieve because the research takes longer, however they feel the pressure to match the higher output. When an academic speaker is introduced at an event they are introduced in the context of their publications, H-index etc. with publication defining their success.

When asked if social networking/Web 2.0 have impinged on ways of thinking about publishing/dissemination these areas were seen as very new, with most academics not that bothered and not having time to engage. These tools are potentially useful for disseminating findings to the public but are not yet not widely taken up. Concern was expressed that findings disseminated through non-standard channels might not be properly attributed. It was felt that there is a need to develop value around different ways of disseminating research findings particularly in the light of RCUK/REF requirements around impact/knowledge exchange and that we need other metrics for measuring the use of these findings such as web page hits and downloads.

Repositories

None of participants had used or were particularly aware of their University's institutional repositories. They were concerned about how to get academics to deposit in repositories and one researcher asked why someone would ever want to deposit into a repository. None of the participants had deposited an item into their institutional repository and they were not aware of ever having discovered or used an item in an institutional repository. There was discussion about the value of aggregating publication data at a geopolitical level, e.g. Scottish. This was initially dismissed as being of no interest but one researcher pointed out that it could be useful to have Scottish-based datasets (e.g. population data). The RAE/REF are seen as acceptable drivers for creating institutional repositories and the participants would deposit in an institutional repository if required to do so by a senior manager. These researchers do not use Google or Google Scholar for research purposes – they go to sources such as Medline and Web of Science. They were not troubled by the idea of using an author's final manuscript as long as it had gone through the peer review process, but were concerned about how to cite these items. It is imperative to demonstrate the value of repositories if we want researchers to engage with them.

Grey Literature

There was a view that it should be easy to capture this kind of material as much of it is now computer-generated but some concern was expressed over the control of information, such as a datasets, and it was suggested that this material could be listed at metadata level even if academics are not happy about making the full content available. It is helpful to have repositories of material such as PhD and MSc theses, many of which are currently unavailable. There is a lot of methodological work that does not get published and researchers can spend a lot of time replicating work already done. When asked about the role of institutional repositories for hosting datasets the point was made that if these datasets are in distributed repositories the information might be spread too thinly and difficult to find, so why not have a centralised database and a single interface for accessing all this information. Difficulty of getting research published which shows negative results and repository might be a platform for disseminating this kind of research.

Research services
One view was that the Library does not offer anything to researchers any more – there is no need to go for search services, for example, these services are all back on the researcher's desk top (Not necessarily seen as a good thing). New researchers have no memory of Library services and might not seek them out so if a member staff cannot access a journal article online they may only read the abstract and not go to the Library and read the journal article in full. The Library is not promoted properly to researchers and emails from the Library, for example, listing services, are ignored because they are not seen as a priority. One participant pointed out that there is one Liaison Librarian for whole of her College. This Librarian is excellent but far too busy, however researchers do value that support when they get it and would welcome more individualised support. Information professionals are being appointed at departmental level undertaking such activity such as supporting grant applications, dissemination of publications and pulling together evidence for researchers.

20th May 2010

Participants

i) Three members of library staff
ii) Anthropologist working in public health. All work done on an electronic archive of 70 million digitised pages from tobacco industry lawsuits.
iii) Lecturer in archaeology.

Pressure to publish
There is constant pressure to publish something different and new but there are constraints over word counts and other data that can be included in articles. The push to publish in mainstream journals is militating against dissemination of research in– for example one researcher prefers to publish in an open access digital resource for her subject, as she can get feedback from colleagues all over the world. Her motivation is to disseminate the research as widely as possible and to be able to read research from all over the world however she is under pressure to publish in journals that are expensive and less accessible, as the University requires her to have three high quality publications between 2010 and 2012. Given the time lags over publication, in practice articles need to be ready by the end of 2010. This researcher feels the pressure to publish is starting to influence her research direction and research profile. The competition to publish is leading to impact factors driving where researchers are placing publications, and this is having a trickle down effect on getting published, how long it takes to for articles to appear and the amount of time spent on peer review etc. The REF is creating a huge amount of pressure to publish and to publish in particular journals. It is important for universities to keep their public audience and part of that is publishing in popular journals and publishing popular books. The impact agenda is pushing universities to make links with public bodies such as national libraries, museums and others, particularly in the arts and social sciences, which can be beneficial to both parties. There is a model of publication in medical sciences which is about collaboration and multi-authored papers and this model is being adopted within the arts and social sciences. One university policy is that authors should be first, second or last on a paper.

Grey literature
In subjects where there may be a requirement for reporting, institutional repositories might be somewhere to deposit interim reports, findings of small research projects etc. The institutional repository is a better place to deposit this kind of material than project websites or material on websites which might disappear. In pre-digital days part of a research archive might include letters negotiating access to archaeological sites, for example, which form part of our social history. There is little interest in emails and other forms of digital interaction such as blogs and wikis being preserved in this way. Researchers might be interested in institutions archiving this kind of material. Are we poorly positioning repositories in the research process and could repositories be used to preserve this kind of material? There may also be a potential role for repositories in the curation and preservation of archives of, for example, eminent staff and other supporting research materials rather than simply recording published output.

Repositories
A researcher's first instinct is to go to a subject repository, or whatever the best way is into finding out about that subject, and therefore institutional repositories need to offer subject access. Researchers will share research with other on an informal basis rather than risk the early 'publishing' of research

which might prevent its formal publication. One researcher commented that it would be irrelevant to create an archive based on geo-political organisation as researchers are not necessarily working that way, but another researcher felt that a global list of Scottish Archives could be very useful, particularly for identifying who is working in which fields of research, what data they have collected and avoiding the duplication of others' research.

Research Services
One researcher commented that the Library in his institution is 'user-unfriendly' and hostile and that he has had little contact with liaison librarians. A researcher has deposited data in the Archaeological Data Service, she uses a service called OASIS to create the metadata for her datasets and she lodges multiple digital copies of this data (for example to the Regional Council) and a copy could also be deposited into an institutional data archive. Participants were asked about the idea of 'open scholarship' i.e. dynamically connecting with other researchers while they are doing their research – it was pointed out that you need an institutional culture which promotes this kind of behaviour and it is also very important to include non-University research bodies in discussions about open access and open scholarship.
Do researchers need more assistance? One participant stated that while information specialists can be very helpful, but for real specialist help he would go to another researcher. Another participant said that she feels she is unusual in widely publishing digitally and using digital archives – she feels that, on the whole, researchers are much more interested in using digital materials than in depositing their own materials.

Appendix C: New Modes of Research Dissemination and Support

'The diffuse knowledge that is embedded within and suffused throughout every university is a form of local content that most institutions have barely begun to tap. Energizing and leveraging this largely latent capacity is critical to the academy's future.'2
(Dan Hazen, Associate Librarian of Harvard College for Collection Development)

Introduction
In the course of 2010, with colleagues at the University of Edinburgh, we undertook two surveys to look into new modes of research dissemination in universities. The first of these was a survey conducted within a group of Scottish Higher Education institutions as part of the work of the JISC-funded ERIS (Enhanced Repository Infrastructure for Scotland) project. [3] Three hundred researchers were surveyed from across Scotland's 19 higher education institutions. These are:

University of Abertay, Dundee†
University of Aberdeen
University of Dundee
Edinburgh College of Art*
Edinburgh Napier University†
University of Edinburgh†
Glasgow Caledonian University
Glasgow School of Art*
University of Glasgow†
Heriot-Watt University*
University of the Highlands and Islands Millennium Institute
Queen Margaret University†
Robert Gordon University
Royal Scottish Academy of Music and Drama*
Scottish Agricultural College*
University of St Andrews

[2] Hazen, Dan (June 2009) 'Rethinking Collections in the Harvard College Library: A Policy Framework for Straitened Times, and Beyond: Discussion Paper and Action Plan http://hcl.harvard.edu/collections/hcl_collections_content_strategy.pdf
[3] http://eriscotland.wordpress.com/

University of Stirling†
University of Strathclyde†
University of the West of Scotland*

* *Institutions without an institutional repository at the time of the survey*
† *Institutions with a deposit mandate at the time of the survey*

They were asked about their experiences of and attitudes to repositories in general, with a particular focus on their role in research dissemination. This survey formed part of Work Package 1 of the ERIS Project 'Enhancing Researchers' Engagement with Repositories'. In its other work packages, this project is looking at 'Enhancing Curation and Preservation Processes within Institutions', 'Technological Enhancements for Improved Research-centric Functionality and Technical Synergy with the Institutions' and 'Developing a … Policy Framework for Organisational and Financial Sustainability'.

An important factor to bear in mind in analysing the responses from these researchers in Scotland is the role played in the UK by the Research Excellence Framework (REF – formerly known as the Research Assessment Exercise). We described the UK's approach to research assessment in a report published in 2009:

> Research assessment has a long history in the UK. There have been six national research assessments exercises since 1986. The most recent was concluded in 2008 and included an assessment of 2,344 submissions from 159 Higher Education Institutions … Higher Education Institutions take the RAE seriously because the results are influential in the selective distribution of research funding by the various higher education funding bodies.[4]

The last Research Assessment Exercise decided how £1.6b of government funding per year would be allocated across the UK's universities. At this level of funding, universities take the exercise very seriously indeed, and all of their researchers are drawn in to the effort to identify their best recent publications, and to ensure – usually with the library's help - that the publication data is correct, and that each individual's research profile is maximised in order that the university can score the highest possible mark in the assessment. It is with this very dominant assessment environment as context, therefore, that we should consider this data.

The second survey was undertaken with the help of a Working Group drawn from a subset of RLG Partnership institutions. This project sought to survey subject librarians[5] in research libraries from a number of different countries, to find out how much their jobs had changed in recent years with the advent of new modes of research dissemination. Ninety-two subject librarians were surveyed from the following countries: Australia, UK, Finland, Germany, Greece, Ireland, Netherlands, Portugal and Spain.

Methodology

Responses in both surveys were gathered by means of web-based questionnaire. They have been aggregated into five identical disciplinary groups across both surveys – Arts & Humanities, Biosciences, Clinical/Health, Science & Engineering and Social Sciences. In the case of the survey of subject librarians there were a number of respondents whose responsibility applied across more than one of these disciplinary groups. In these cases, they were assigned to the group that they gave as the first on their list of subjects.

Both survey projects were coordinated by James Toon of Edinburgh University Library, who was Project Manager for the ERIS Project, and also Chair of the RLG Partnership Working Group on Supporting Research Dissemination. Selective results from both surveys are presented below. We begin with a report of the key findings from both surveys. This makes reference to the data which is presented in tabular form in the data section of this report. We have limited the data presentation to those elements which are most relevant to our findings in the area of research dissemination and its support. Some of the data gathered has therefore been omitted below, but the full dataset is available upon request. The tables showing response percentages are presented in order of size of response

[4] Key Perspectives Ltd. 2009. *A Comparative Review of Research Assessment Regimes in Five Countries and the Role of Libraries in the Research Assessment Process*. Report commissioned by OCLC Research. Published online at:http://www.oclc.org/research/publications/library/2009/2009-09.pdf

[5] 'Subject librarians' is used in a general sense throughout this report, to encompass the various designations in use, including 'Academic Liaison Librarians', 'Information Consultants' and others.

by subject group. Some of the question wordings in the table captions have been amended from the original survey questions to aid reporting clarity.

Distribution of Samples

Survey invitations were sent out widely, and there was little scope within either project to manage the responses for even distribution across subject groups. The relative distributions of responses are somewhat different across the two surveys. The survey of researchers has a more even distribution across the subject groups, from Social Sciences (29% of total) to Clinical/Health (14% of total). The responses from subject librarians ranged more widely, from 35% for Arts & Humanities, to just 9% from Biosciences. However, the distortion that this may imply is offset to a degree by the fact that several respondents in this latter survey had responsibilities across more than one subject group. Also, subject librarians are inevitably more generalists than are academic researchers, and so the subject group into which they are placed is of less significance than is the case in Figure 1.

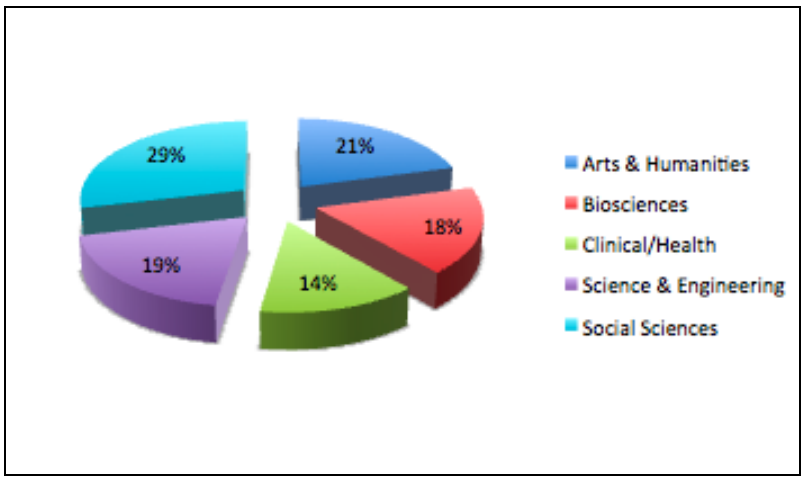


Figure 1: Survey of Researchers in Scottish HE institutions: Respondents by Subject Group

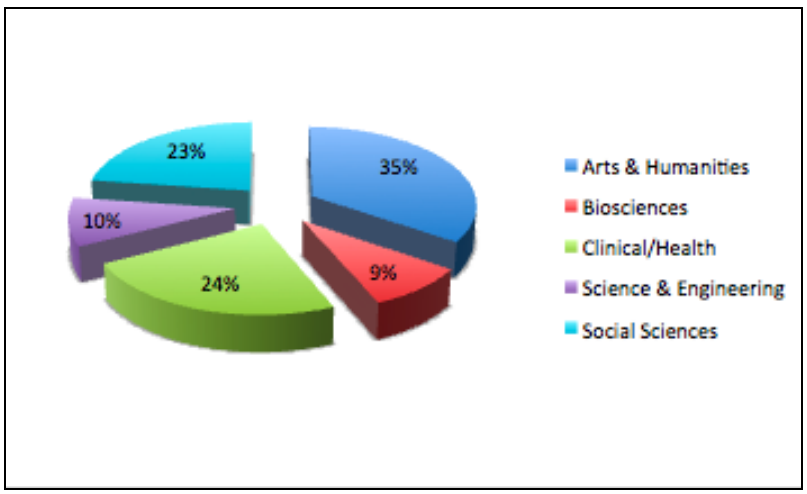


Figure 2: Survey of Subject Librarians from Research Libraries in Nine Countries: Respondents by Subject Group

Key Findings

Survey of Researchers

New Territory for Libraries

In considering new modes of research dissemination in these surveys, we were particularly interested in understanding the scope for library roles in relation to what is sometimes described as the 'inside out' function, as opposed to libraries' traditional responsibility of managing the 'outside in'[6] – ie publisher-produced material in journal or monograph format. The most obvious alternative vehicle for this type of content in recent years has been the repository, and where libraries are most concerned, this means the institutional repository, which is normally managed by the institutional library. Many of the questions in both surveys therefore focused on the institutional repository, although other new forms of dissemination also feature.

The Institutional Repository

We first asked the researchers if they were aware of their institutional repository – whether one existed, and what services it offered (Table 1). The response rates were in the range 59% (Clinical/Health researchers) to 74% (Social Sciences). This might seem a little low, given library advocacy activities going back some 10 years or so, but with some of the smaller and newer institutions having only recently launched repositories, it is probably a reasonable figure. Six of the institutions surveyed had no institutional repository at the time of the survey (see list on page 2 above). Respondents from those institutions were still able to answer all but one of the questions, and they were deliberately included. However, their responses to the next question were removed from the analysis. This asked whether the respondents – or someone acting on their behalf – had ever deposited in an institutional repository (Table 2). Inevitably, the percentage figures dropped at this point. Only the Science & Engineering group scores above 50% here, with the average figure for those who have deposited in their repository standing at 44%.

Deposit mandates (policies requiring that qualifying outputs be deposited in the institutional repository) had in a few cases been introduced to institutions before our survey took place. In some cases these were clearly only at very early stages. These were 'institutional mandates' as defined in the ROARMAP (Registry of Open Access Repository Material Archiving Policies)[7]. Nonetheless, we thought it worth looking at the rates of deposit in these repositories compared to that in institutions without mandates (Table 3). The results seem to give a very positive reinforcement to the use of mandates to increase deposit, with the average in mandating institutions being 58%, compared to 46% in those without mandates. It is difficult to know quite how to interpret this. It is possible that many of the respondents had not produced any qualifying outputs since the mandates had been introduced, and so this figure should be considered an under-estimate. At any rate it does seem to be the case that mandates will increase deposit rates, but how close to 100% is possible will require to be calculated once their operation has been given time to be embedded in cultures and workflows.

Scientists, at 43%, had the highest percentage of alternative sites (their own web sites and pages) to host their research publications. Clinical/Health researchers were least likely to have such an alternative (Table 4). We asked about subject repositories. Those who use them are enthusiastic about them, and vaunt their superiority to institutional repositories. However, the figures for those using them were lower than expected. Scientists and bioscientists were the highest users, at 18%-19%, while, at the lowest (Clinical/Health), the figure drops to 10% (Table 5).

We sought opinions on the benefits of institutional repositories (Table 6). The Clinical/Health group produced the highest figure in support of the view that institutional repositories give the research output of the institution more exposure (88%). Biosciences and Science & Engineering, both at 68%,

[6] See for example Dempsey, L 'Outside-in and inside-out', January 11 2010 http://orweblog.oclc.org/archives/002047.html

[7] http://www.eprints.org/openaccess/policysignup/

were the least convinced. This contrasts interestingly with the ranking for the use of subject repositories in Table 5. Clinical/Health researchers, with perhaps a stronger allegiance to their institution than other groups, are most prepared to back the value of the institutional repository, whereas the scientists and bioscientists are most interested in repositories that belong to their disciplines, putting the institution's services at a lower priority.

We asked them how repository deposit could be made easier. The main theme of the responses (Table 7) was *'take the work away from me'*. In several cases, clearly this already happens, with identified staff in the library or research administration doing the deposit work. This would suggest that any ambitions either of these services might have to transition the work onto researchers directly at a later date are unlikely to be successful. One respondent felt that the support services were not showing enough courtesy or gratitude to the researchers who were, it is implied, doing their work for them.

Several respondents also pointed to the 'flakiness' of the software. The majority of these latter remarks were from researchers in the biosciences, who may have been contrasting institutional repositories with subject-based repositories they used such as UK PubMed Central. There was also some criticism of the lack of author name disambiguation. Since researchers are familiar with using their library catalogues, there may be surprise that the same standards are not in place in this new library-managed database.

One researcher had moved institutions and lamented the fact that their repository deposits had not moved with them. This points to a requirement for a sophisticated archive of material that would require join-up across universities in a way that is not possible with the current infrastructure.

We asked them (Tables 8-10) what sort of material they thought should be contained in an institutional repository, and gave them three choices. There was an interesting disciplinary division of responses here. Arts & Humanities academics were the most keen on limiting the institutional repository to peer-reviewed material. But they were the least prepared to accept a mandate (13%, where the average was 29%). Encouragingly, the 'liberal' view of a repository got the highest average response (37%), with scientists and social scientists most prepared to see the institutional repository accept non-peer-reviewed material alongside peer-reviewed. From a library perspective, this is good news for the development of 'inside out' stewardship and for the idea that the institutional repository should do more than just serve the Research Excellence Framework process. But it might also reflect a view among scientists and social scientists that the traditional publisher-controlled journal system should be preserved as the authoritative expression of the peer-reviewed literature, and therefore that the institutional repository is part of a quite distinct publication service. Many librarians would take issue with this, since it assumes no eventual transfer of publishing behaviour to the 'Green' Open Access model (ie Open Access publishing based upon self-archiving rather than via publisher-produced Open Access journals), which has the potential to subvert the commercial publisher-controlled system and bring cost-efficiencies to bear.

Citations and Impact

Researchers were asked whether, after publishing their research, they actively monitored the impact of their work through evidence of citation (Table 11). The range in response to this question was large – with Biosciences at one end (70%) and Arts & Humanities at the other (29%). Citation-based impact measures are growing in use and importance as research productivity intensifies and becomes more competitive. They are of course more readily available for the journal literature – which is central to publishing activity in the sciences and medicine - than for the book-based literature, which is used much more by the Arts and Humanities.

We then asked whether academics would value new services based upon citations. (Table 12). In general, the sense is that academics would not trust a service provided institutionally. They prefer those that are provided by the publishers or other services within their discipline, or the large multidisciplinary services (Scopus and Google Scholar are mentioned a few times). There are a few calls for institutionally-provided services, but they are not convincing. Yet expertise from the library on what is available and how to make use of it does seem to be something that would be welcomed. In other words, libraries might find academics more interested in being trained in how to use the existing

tools of their disciplines – and others they might not be aware of – than in new services developed *de novo* within their libraries. Given their strong sense of urgency around their own work, and need to be as efficient as possible with their time, libraries should heed this message.

'Hidden' Outputs

In the next group of questions we sought to find out the scale of the 'hidden' outputs of research – the material that could be of use to researchers more generally if there was a way to expose it. We asked them whether they had any scholarly or scientific work related to their research that could be valuable to other academics, but for a range of reasons is not published. This question (Table 13) elicited a very uniform response (in the range 50%-60%) across all disciplines – and a majority in each discipline see value in 'hidden' research materials that they produce. When we then went on to ask them whether they thought other researchers in their field had hidden work that would be of value to them, the majorities were considerably greater (Table 14), though there is more disciplinary variation. Scientists are the most tentative, perhaps fearing the information overload/data deluge implications. But it is clear that there is a mass of unstewarded 'inside out' material in universities – and this surely represents a new professional front for academic libraries. Why do more academics think that their peers have valuable hidden material than claim to have it themselves? This is likely to be for the same reason, identified in various studies, that they want access to other researchers' data, but are less keen to provide that same access to others. This was found in the JISC-funded StORe project in the UK, for example:

> As observed by a social sciences respondent, researchers' attitudes to enabling access to their data will depend to a large extent on whether they are behaving as producers or users of data, with producers concerned to protect their endeavours from predatory access to source repositories by their competitors.[8]

We went on to ask whether they do anything to make this unpublished material available via a local or a central repository service. The responses are shown in Table 15. Time, again, features largely. Some researchers have the best of intentions towards such material, but can't spare the time from their research or teaching schedules to work on it. Two respondents mentioned their aspirations to make the material available for use after they had died. In general, practice is chaotic, and misinformation and misunderstanding abounds around copyright, intellectual property and both digital and data preservation and curation. Data has to be considered separately from presented findings (ie versions of journal articles, reports or working papers) – and openness in the case of data can justifiably be delayed, or embargoed. An interesting point was made about null results, which are important research findings that are often hard to get published. Can the institutional repository perhaps offer a 'backstop' publishing service for such material?

They were then asked (Table 16) whether proprietary or commercial reasons affected the distribution of their unpublished work. At the top of the range of those who replied in the affirmative were Arts & Humanities and Science & Engineering. In both cases a quarter of respondents said their unpublished work would be restricted. The responses (Table 17) reveal a mix of valid reasons for restriction, together with simple lack of knowledge which generally creates risk-averse behaviour – and which it should be in the best interests of libraries to tackle. We asked (Table 18) about other reasons for not making their unpublished work available. The most common theme to emerge in the responses was *'lack of time'*. This means lack of time both to learn about good practice, and lack of time to conform to good practice. Competition also emerged, at times a little sheepishly, acknowledging that the world of research reputation in the UK can be cut-throat and not necessarily in the best interests of scholarship. Risk-aversion also appears; academics live and die by their reputations. Inferior work published on the web could jeopardise careers. One very honest comment pointed out that errors in the methodology would be more easily exposed. Others suggest a 'once bitten, twice shy' position – or, in other words, *I did the noble thing once and got ripped off.*

Indeed there seems to be a boundary line that academics are aware of, on one side of which there are outputs that they don't want to 'get out there'. Sometimes this is for reasons of quality; other times it is because these outputs are about learning and engagement with learners, rather than material fit for the 'research corpus'. The idea of a mandate threatens to blur the distinction between these, and creates the risk that a researcher's reputation could be negatively affected if some of this other work is not protected. So libraries might do well to consider how to channel their efforts into developing the

[8] Pryor, Graham (2007) 'Attitudes and aspirations in a diverse world: the Project StORe perspective on scientific repositories', *International Journal of Digital Curation* 1 (2) 135-144 http://www.ijdc.net/index.php/ijdc/article/viewFile/32/21

services that respect those boundaries, rather than weighing in behind mandates that threaten to expose the entire spectrum of a researcher's outputs.

One of the most succinct responses was simply *'laziness, apathy, competition'*. The laziness and apathy surely relate to those tasks which are considered an additional burden on top of the competitive pressures that are so predominant. There were also responses which pointed to curatorial failure; there is no infrastructure. This is a challenge to our community. There was some criticism of institutional repositories in response to this question - criticisms which would appear again in response to the next set of questions.

Other Repository Services

Next came a key question (Table 19): *'Are there other services you can think of that repositories could develop to broaden their appeal and usefulness to you?'* It brought forth a number of considered – and in fact very useful and quite knowledgeable - responses, but also unleashed a stream of skepticism about institutional repositories from some respondents – and a corresponding show of support for subject repositories instead. Some were blunt, eg *'If all institutional repositories ceased to exist tomorrow the loss to scholarship would be negligible'.* Others again gestured towards the idea of a scholarly archive, which researchers seem to feel should be bolted on to the current journal system: '*I never use institutional repositories as prefer to obtain the formatted, corrected version of journal articles directly from the publisher. If my institution doesn't subscribe, then I'd contact the authors directly. Pre-prints are harder to read and authors often don't bother to implement the changes suggested by referees or publishers in the archived article.'*

We asked how strongly the Research Excellence Framework exercise in the UK motivated them to publish (Table 20). The Clinical/Health response was the highest here (78%), with Science & Engineering at the bottom, at 50%. A general observation might again be that Clinical/Health academics are more institutionally-oriented than is true in other disciplines – perhaps because of a greater business focus. But of course, in the case of the UK, helping the institution to do well in the REF is important primarily for the financial rewards that this brings to departments themselves. So there is much more to this level of response than altruism.

The next group of selected tables from this survey (Table 21-25) all measure the levels of disagreement with the advantages of an institutional repository as might be claimed by libraries – that they allow researchers to find colleagues' research; that they give the work of the individual researcher more exposure; that they offer a service to researchers who cannot afford to access the toll-gated literature; that they help libraries to move away from the increasing costs of scholarly publishing; and that they are an exciting innovation. Researchers generally assent to these propositions, but the levels of dissent are interesting. Scientists and bioscientists are the most skeptical, with levels of disagreement running at between 30%-40% - though this reduces somewhat in relation to the issue of exposure, and considerably when considering the issue of access for impoverished researchers.

Finally, we asked two questions about the potential disadvantages of repositories (Tables 26-27). First, that they expose more research to plagiarism. The highest level of agreement here is from bioscientists (38%), a group that other studies have shown are particularly worried about their research being 'scooped' (see, for example, the 2007 Ithaka report on scholarly communications in the biosciences).[9] Second, that they may breach the confidentiality of data in some research. Again, the biosciences are quite concerned by this, but the Arts & Humanities academics are the most worried (48%), presumably because of the dangers of working with primary source materials relating to deceased individuals whose estates may be anxious to protect their rights.

Survey of Subject Librarians

Subject Librarian Roles

[9] Quinn, Meredith and Kim, Jennifer (2007) *Scholarly communications in the biosciences discipline: a report commissioned by JSTOR*
http://www.ithaka.org/publications/pdfs/JSTOR%20BioSci%20Study%20Report%20Public%20final1031.pdf

The survey of subject librarians began by asking them how they saw their roles having changed in recent years (Tables 28-30). The highest average figure, at 51%, was for an increase in the amount of learning and teaching support. Research support, at 36%, was next, and collections management (14%) was the least likely element of the role to have increased. While, perhaps predictably, subject librarians in the Arts & Humanities were the most likely to indicate an increase in collections management work, they were the least likely to point to an increase in research support. Librarians in the social sciences saw the greatest increase here (47%). But a general conclusion must be that increased research support is not keeping up with support for learning and teaching. Given the requirements implied by the survey of researchers above, this may be a cause for concern. There was also scope for respondents to mention other ways in which they felt their roles had changed (Table 31).

New Mode Activity

The next set of tables (Tables 32-42) asked them to identify how often they were engaged in support for new forms of research dissemination, including the use of the repository. Here we report on those activities that they engage in regularly – ie at least a few times per week. The numbers are low. There is little evidence of them spending much time advising academic staff about the institutional repository, Open Access, bibliometrics, or on making improvements to the repository, and virtually no activity in respect of the management of research data. The figures are higher in relation to activities that may also have occurred in pre-repository and pre-web 2.0 days, such as the provision of advice on copyright and intellectual property, and participation in procuring and developing new tools and services for research. They also rise a little in relation to librarians' own use of web 2.0 tools to 'market' research. Part of the reason for these generally low levels of repository-related work are no doubt due to confusion over whether this area of activity should belong with subject librarians, or with other service sections. This comes up more clearly later in the survey.

We asked about traditional subject librarian tasks that had now been superseded (Table 43). Not surprisingly, activities related to print publications came up a few times, as did reference work. We also asked (Table 44) whether their changed roles were being recognized in adjusted job descriptions. Fifty seven percent said 'No', whereas 84% thought they should be (Table 46). Where changes had been made, these were said to include reallocation of more traditional skills to lower grades of staff, and 'repository work' (Table 45). Of the 16% who thought that these new tasks should not be part of their remits, a range of alternative services both within and outwith the library were listed as being more appropriate providers (Table 47).

Nonetheless, 67% of the total group surveyed thought that web 2.0 tools and approaches were appropriate in supporting research dissemination (Table 48), 84% had found them a positive experience (Table 53), and 47% said they personally made use of them (Table 49), with blogs being the most-cited example (Table 50). A small majority also agreed overall (Table 51) with the idea that subject-focused social networks were an important area for subject librarians to be involved in research dissemination, and a few of them gave examples of social networking initiatives that they had undertaken (Table 52).

Conclusion

A general conclusion might be that subject librarians are reasonably keen on adopting new methods of supporting their researchers, but few of them are involved in supporting the adoption of these methods by the researchers themselves. There appears to be a slow and uncertain transition towards the development of what Jim Neal of Columbia University Library has called 'Subject Librarian 2.0'[10].

Meanwhile, the researchers themselves have already adopted many new methods of research dissemination, but they do not see the institutional repository as particularly important here. They also

[10] Williams, Karen (2009) 'A framework for articulating new library roles' *Research library issues: a bimonthly report from ARL, CNI, and SPARC (August 2009) http://www.arl.org/bm~doc/rli-265-williams.pdf*

tend in general to feel that the traditional form of research publishing in journals should be left as is, and new forms should be added to it in ways that are at least parallel, or complementary.

Researchers have considerable needs, which libraries – institutionally and collaboratively – are not yet meeting. Libraries need to give some attention to the question of a scholarly archive that spans the peer-reviewed and the 'grey' corpus and will require significant infrastructural join-up by a number of players. They also need to think about the best way to deploy their subject librarians in ways that make use of their subject knowledge but also go beyond it, to embrace knowledge of subject communities, and how they now behave in research universities.

Response Data: Scottish Researcher Survey

Social Sciences	74%
Science & Engineering	73%
Arts & Humanities	71%
Biosciences	60%
Clinical/Health	59%

Table 1: 'Are you aware if your institution has … a repository (or set of repositories) and what associated services, if any, are offered?' - 'Yes'

Science & Engineering	56%
Social Sciences	49%
Clinical/Health	44%
Arts & Humanities	42%
Biosciences	30%
Average deposit rate	*44%*

Table 2: 'Have you, or someone nominated by you, ever deposited items into your institution's managed repository?' - 'Yes'[11]

Average deposit rate for institutions with deposit mandates	58%
Average deposit rate for institutions without mandates	46%

Table 3: Average deposit rates in institutions with mandates compared to those without

Science & Engineering	43%
Social Sciences	32%
Biosciences	28%
Arts & Humanities	19%
Clinical/Health	10%

Table 4: 'Do you also maintain your own web site or repository to make copies of your research publications available on the web?' - 'Yes'

Biosciences	19%
Science & Engineering	18%
Social Sciences	15%
Arts & Humanities	11%
Clinical/Health	5%

Table 5: 'Do you also deposit full text copies or links to your research publications in independent subject-based repositories?' - 'Yes'

Clinical/Health	88%
Social Sciences	74%
Arts & Humanities	68%
Biosciences	68%
Science & Engineering	68%

Table 6: 'Institutional repositories give the research output of the institution more exposure' - 'Agree'

[11] The responses to this question have been analyzed after first removing those from the six institutions with no institutional repository.

- *It would be useful to make it easier to add information about publication to a repository without having to worry about uploading the file (i.e. add a full reference but not submit the file). Furthermore, due to the complexity of copyright for some publishers, and the fact that we are instructed to upload all publications, I believe it is the responsibility of the institution to enquire about permissions and ultimately decide which version of the publication (if any) should be uploaded. In my experience, it has fallen to me to chase publishers regarding copyright permission.* (Arts & Humanities)
- *Clearer guidelines about whose responsibility it is to deposit it (e.g. own, department admin staff, school/faculty admin staff).* (Arts & Humanities)
- *Yes - it is difficult software, with gremlins.* (Arts & Humanities)
- *Yes: have a copy paste option instead of typing everything.* (Arts & Humanities)
- *If I could send material by email attachment to someone else to do it for me!! I'm not opposed to doing this in the case, particularly, of journal articles or books but I could do with a little more technical help/training and TIME!* (Arts & Humanities)
- *It is clunky when handling multi-author works.* (Biosciences)
- *It is a crude system (online) and does not work properly* (Biosciences)
- *The database is awkward and requires filling individual boxes rather than copying or uploading a standard citation list. Passwords need to be remembered and the cite is vulnerable to frequent crashes.* (Biosciences)
- *We forward to a named person.* (Clinical/Health)
- *It goes to REF administrator.* (Clinical/Health)
- *It would be nice if the various details such as issue no, volume, page no etc, as well as the abstract could be extruded automatically from the DOI and be filed into the institutional repository form without me doing this very tedious work.* (Clinical/Health)
- *Some staff members seem to appear under several names - e.g. S. Rogers, Simon Rogers etc.* (Science & Engineering)
- *Easy for me; it is done by secretarial staff.* (Science & Engineering)
- *Still getting used to it and it seems overly time consuming to do. Different systems in place for recording publications at Faculty and departmental level lead to lots of duplication.* (Science & Engineering)
- *I need assistance to produce the version of my publications that can be put on the repository. It looks quite onerous, but support has been good.* (Social Sciences)
- *Less bureaucratic; no compulsory; more courtesy and gratitude.* (Social Sciences)
- *I send the material to the library staff and they process this for me.* (Social Sciences)
- *The categorisation system is cumbersome and time consuming. The time taken for input items to be reviewed by the library (who administer the system) and finally deposited. The page where input items 'lie' until fully accepted into the system is not at all user friendly.* (Social Sciences)
- *Not using pre-internet classification system (LoC); speeding up process of acceptance of submissions.* (Social Sciences)
- *I have just moved institutions and am currently exploring the process of depositing my work in the Strathclyde repository. It would be easier if there was a straightforward way that my deposits in the repository of my previous institution could just be transferred over. It would also be easier if there was an individual or team that would do this work on my behalf (this may be the case - I am just exploring the mechanisms available to have things put into the repository here).* (Social Sciences)

Table 7: 'How could the repository deposit process be made easier?' – Comments

Arts & Humanities	52%
Biosciences	36%
Clinical/Health	29%
Social Sciences	26%
Science & Engineering	23%
Average	*33%*

Table 8: 'What should be contained in the institutional repository?' – 'Only peer-reviewed conference papers, journal articles and theses that staff choose to deposit'

Clinical/Health	37%
Science & Engineering	36%
Social Sciences	32%
Biosciences	28%
Arts & Humanities	13%
Average	*29%*

Table 9: 'What should be contained in the institutional repository?' – 'All the peer-reviewed outputs of the entire institution compulsorily'

Science & Engineering	41%
Social Sciences	41%
Biosciences	36%
Arts & Humanities	35%
Clinical/Health	34%
Average	*37%*

Table 10: 'What should be contained in the institutional repository?' – 'Anything (non-peer-reviewed papers and drafts) that academic staff wish to deposit'

Biosciences	70%
Science & Engineering	61%
Social Sciences	49%
Clinical/Health	46%
Arts & Humanities	29%

Table 11: 'Once published, do you actively monitor the impact of your work through evidence of citation?' - 'Yes'

- *I don't know how to monitor citations.* (Arts & Humanities)
- *Yes, I would like more information; also I often forget to check, but I do not know what the institution could do about this.* (Arts & Humanities)
- *I have no idea how to do this reliably. All I know how to use is SSCI and Google Scholar searches.* (Arts & Humanities)
- *This is difficult to implement in Humanities. Being book-based and traditionally more conservative, quotations are often made only in paper medium, thus difficult to find (unless one knows it is there). Usually one finds out the level of quotation by chance or by networking.* (Arts & Humanities)
- *My monitoring doesn't amount to much more than checking indexes and bibliographies of new publications. But I do do this!* (Arts & Humanities)
- *Informally. Could do with working out exactly how to use Google Scholar. I'm less good than I should be at tracking citations.* (Biosciences)
- *No - leave it to the people who are good at it.* (Biosciences)
- *Existing processes are fine, I think.* (Biosciences)
- *Difficult to get all citation data in the one place eg Web of Science doesn't include some open access journals, or review journals - Google Scholar can help, but coverage also limited. Institutional access to Scopus might help?* (Biosciences)
- *Science citations are very easily accessed; no further action required.* (Biosciences)
- *Yes. Publicise it and send a an update to managers and researchers.* (Clinical/Health)
- *Would be great if they could produce and send me a list of references with citations of those works on a regular basis.* (Clinical/Health)
- *Ideally, some monitoring of citations would happen at institutional level.* (Clinical/Health)
- *Have done this periodically but have not found it an easy process.* (Clinical/Health)
- *If this could be linked to the repository so I could easily search all my papers in one go it would be great.* (Clinical/Health)
- *Maybe collect the info centrally and send us updates - but it is not really a problem.* (Clinical/Health)

- *The information provided by publishers such as the BMC journals regarding the number of times a paper is accessed is valuable in the first year before citations begin - citations have long lag.* (Clinical/Health)
- *I would like a clearer, easier mechanism for doing this.* (Science & Engineering)
- *Not sure if this is something that institutions could do, but the new PLoS article metrics are very interesting.* (Science & Engineering)
- *I don't know what existing automatic processes are.* (Science & Engineering)
- *I used to when we had access to Scopus.* (Science & Engineering)
- *Though I occasionally do search for papers citing me, I don't think it's something my university should be putting money or resources into.* (Science & Engineering)
- *It would be useful for university to maintain a citation register.* (Social Sciences)

Table 12: 'Is there anything your institution could do to provide more information on citations to your published works?' - Comments

Arts & Humanities	58%
Clinical/Health	56%
Social Sciences	55%
Biosciences	51%
Science & Engineering	50%

Table 13: 'Do you have any scholarly or scientific work related to your research that could be valuable to other academics, but for whatever reason is not published?' - 'Yes'

Clinical/Health	90%
Social Sciences	89%
Biosciences	87%
Arts & Humanities	86%
Science & Engineering	70%

Table 14: 'Do you expect that other academics in your field have such unpublished work that would be valuable to you and others in your area?' – 'Yes'

- *I am organizing field notes for use after my death.* (Arts & Humanities)
- *This material is, in my view, primarily of post-mortem interest.* (Arts & Humanities)
- *I make sure that I copy everything to a departmentally shared network and encourage all my colleagues to do so.* (Arts & Humanities)
- *No. I need to write it up for peer review and submit it to a journal. Preprint repository is not useful to my subject area.* (Arts & Humanities)
- *Some conference papers; I plan to put these on the website of the AHRC research network which I lead.* (Arts & Humanities)
- *I have not done anything to publish the material. I am waiting for copyright to expire.* (Arts & Humanities)
- *No, I do not make this unpublished material available (if I did this would potentially reduce my capacity to publish it/material relating to it in future); I have little material which I would exclude using altogether in future publications, and thus to protect my IP and my chances of publishing such work at a later date I don't tend to make it more widely available after its original dissemination.* (Arts & Humanities)
- *No, but people can find my email address quite easily via Google. I am always prepared to answer specific questions and send what data I have.* (Arts & Humanities)
- *The deposited material should have a DOI so that the work can be referenced and there is evidence of publication date.* (Arts & Humanities)
- *I keep meaning to put it on my web page but can't find an obvious place for it.* (Arts & Humanities)
- *I will leave it to the University Library, but keep it to work on meanwhile.* (Arts & Humanities)
- *I only consider my pilot data unavailable for publication prior to grant submission, as I cannot afford for other workers to capitalise on my efforts at that stage of the project's development.* (Biosciences)
- *Problem is time taken to write up - takes time whether published conventionally or not.* (Biosciences)
- *Selected materials could be deposited, but I would not wish them to be accessed without permission being sought (and granted) in advance.* (Biosciences)
- *No I do not think it appropriate to make such data available. It could be misinterpreted in others hands and I think this is potentially harmful.* (Biosciences)

- *This is related to the problem that null result papers are not often published. I have papers written with null results but have failed to get them published. Null results are just as useful to researchers as significant results.* (Clinical/Health)
- *Not really - I personally don't want to spend time looking at anything less than highly polished final outputs - so would only very rarely look at other people's lectures/reports / unpublished work. my priority is to work up my unpublished work for publication - but time is always the problem.* (Clinical/Health)
- *A significant amount of our work is funded by industry and is commercially sensitive. This often leads to delays of several years in publishing -- and sometimes we ultimately fail to write it up for publication because of the demands of new programmes that commence subsequently.* (Science & Engineering)
- *In the field of education conference papers are sometimes made available through a national database which I think is called the British Education Index. I personally am reluctant to have too many conference papers available, as I usually work them up for official publication.* (Social Sciences)
- *I assess on a one to one bases whether to provide individual access to my Ph.D. thesis. A note of my email address for contacting me is placed on the university's repository, when access to my thesis is requested.* (Social Sciences)
- *Will not place it into repository as this is frowned upon by many social science journals and any future attempts to use such material in papers for publication will render the work to be deemed 'unoriginal'.* (Social Sciences)
- *I have datasets already analysed but that have the potential to be analysed further. I would not be interested in publishing them/making them available to others as it is likely I will carry out further analyses myself.* (Social Sciences)
- *Mostly papers waiting to be written, held up due to pressures of other activity.* (Social Sciences)
- *I have lots of data but lack of time to write it up. This is a workload issue.* (Social Sciences)

Table 15: 'Do you do anything to make this unpublished material available via a local or central repository service, and if so, what?' – Comments

Arts & Humanities	25%
Science & Engineering	25%
Clinical/Health	20%
Social Sciences	20%
Biosciences	17%

Table 16: 'Are there proprietary or commercial reasons that your unpublished work would be subject to distribution restrictions?' – 'Yes'

- *Original research on manuscripts currently under copyright protection and within the purview of a literary estate hostile to scholarship.* (Arts & Humanities)
- *I have little knowledge about the proprietary or commercial aspects of my work. I'm just a scholar.* (Arts & Humanities)
- *I may wish to publish some of it at some point.* (Arts & Humanities)
- *It is my intellectual property.* (Arts & Humanities)
- *Authors' personal papers still in copyright.* (Arts & Humanities)
- *Not formally, at least, but I do wonder about intellectual copyright/ideas not yet under contract with a publisher that a bigger name might be able to take forward more quickly than I can.* (Arts & Humanities)
- *Some of my output is subject to confidentiality: arrangements with the (industrial) funder.* (Biosciences)
- *The view that if work is unpublished then it 'doesn't count' in terms of having been your work - if someone repeated it and published then they would get the credit.* (Biosciences)
- *My data are sensitive and much of my data are under proprietary rights by our funders, a Biotech company.* (Biosciences)
- *A 'not sure' option is missing here.* (Clinical/Health)
- *If we would like to file for patent, publishing in any form is hindering a patent application.* (Clinical/Health)
- *Perhaps ethical reasons if the patient data involved were gathered for a particular purpose and other uses are not covered by the patient consent given.* (Clinical/Health)
- *Commercial exploitation rights for industrially funded work.* (Science & Engineering)
- *I'm not sure, but I don't think we could publish the project documents until the project is finished as we are under contract.* (Science & Engineering)
- *Oral history material that I did not gain consent to store in an archive or repository after I had used it.* (Social Sciences)

- *No, but there is the academic point that I rather have publications than conference papers available, also because citation of conference papers is invisible.* (Social Sciences)
- *Not really, but depends on funders' contract* (Social Sciences)
- *We aim to commercialise part of the output from my thesis. Should this become public it could be commercialised by anybody else.* (Social Sciences)
- *Some clinical based data has to be destroyed after a certain timescale.* (Social Sciences)

Table 17: 'What are the proprietary or commercial reasons that your unpublished work would be subject to distribution restrictions?' - Comments

- *I'm not dead yet, so I may still use it.* (Arts & Humanities)
- *I would have to seek and probably pay for copyright for images. I'm not sure this is worth doing unless a publication is involved.* (Arts & Humanities)
- *Not yet at a stage to merit dissemination - that is why it is not yet published!* (Arts & Humanities)
- *Material which is unpublished but pending publication cannot be made available via depositories etc in order to protect the copyright of the journals/monographs/other publications where it is due to be published subsequently.* (Arts & Humanities)
- *The pile of research on which I intend to base the next 3 years work will remain private till I've done it. The pile of research on which I based the last 10 years work is organised so that I can find what I'm looking for but it would take quite a lot of work to turn it into a tidy database.* (Arts & Humanities)
- *REF submission (unfortunately) works as a disincentive to collaboration across institutions; I wouldn't want my research to be used to promote researchers at competing institutions at my expense.* (Arts & Humanities)
- *I might not want to make it available to everyone; it might be better to have it available for comments from subject experts (sort of like an online working paper).* (Arts & Humanities)
- *Potential discovery that the method used was not appropriate to investigate my research question.* (Arts & Humanities)
- *It might be unfinished or not very good and affect your reputation negatively.* (Arts & Humanities)
- *As above - lectures and seminars are part of an on-going conversation. This is more about intellectual engagement and active, on-going thinking than anything I'd particularly want 'out there' with my name on it. I'm not just a proprietary brand, I'm a person who values critical engagement and conversation for its own sake or for the sake of learning - my own or my students/colleagues. Some of this isn't even about writing or recording something but about the engagement itself as energising, enjoyable and productive in a largely non-material sense.* (Arts & Humanities)
- *Because some data-sets are very complex, and without appropriate supporting/contextual information they could be misconstrued or misused.* (Biosciences)
- *If you aim to publish it eventually, someone else could scoop you before that if it was made available.* (Biosciences)
- *Nowhere to put it.* (Biosciences)
- *Lack of time - focus is on outputs for REF.* (Clinical/Health)
- *In the past, my work has been utilised in others' presentations unacknowledged.* (Clinical/Health)
- *Time required to get it into a format suitable for other people to view would not be an efficient use of time when I have papers I need to get published. So this would take a lower priority.* (Clinical/Health)
- *Still working with the data. Not wishing to make accessible till completed all relevant analyses.* (Science & Engineering)
- *For correct citation of the work the hyperlink to it needs to be permanent. Also, publishers of journals need to accept these references as valid, permanent links.* (Science & Engineering)
- *Laziness, apathy, competition.* (Science & Engineering)
- *None other than not having clear mechanisms to enable such grey publication.* (Science & Engineering)
- *I am still working on the problems for my PhD. If other people solve them before me then I have nothing to write about, therefore no PhD and a waste of 3 years.* (Science & Engineering)
- *Internal information exists in a raw state and can sometimes make recommendations which external interests would take issue with. These recommendations may relate to a commercial product, but may contain information that the manufacturer does not wish to be made public. Better safe than sorry in such circumstances.* (Science & Engineering)
- *We publish the important information at an appropriate time in places where it receives appropriate good notice ... that is not an institutional repository.* (Science & Engineering)
- *The main concern is that such products don't generate measurable citations.* (Social Sciences)

- *Information overload! The publication process is a valuable filter on the quantity of material provided and direct contacts between researchers enable adequate exchange of information that is sufficiently developed.* (Social Sciences)
- *It hasn't been subject to peer review; often it is hugely time-consuming to make materials (data sets etc.) suitable for 'others' to use, and there's little to be gained from doing it.* (Social Sciences)

Table 18: 'Can you please tell us about any other reasons that you might have to not make this information available?' - Comments

- *They could do all the data entry and other donkey work for the researchers and let us get on with the research.* (Arts & Humanities)
- *Citation information (including DOI number, ISBN info) for article available in useful formats (BibTex, EndNote). Fixed URL to use to reference paper elsewhere (e.g. personal page).* (Arts & Humanities)
- *Offering a copyright/data protection advice and check service.* (Arts & Humanities)
- *Integration of the material they hold into subject-specific bibliographies or other search tools.* (Arts & Humanities)
- *I want to be able to display links to all articles in the depository that I have written on my own web page. My main concern with repositories is that my experience of searching for other academics' materials is that they are generally easier to find when posted to those individuals' web pages than when in institutional depositories. That is, sometimes these depositories make work LESS visible/available.* (Arts & Humanities)
- *Subject-based repositories are best. See PhilPapers.* (Arts & Humanities)
- *In my subject area it is far more common for people to publish work on personal webpages or subject specific repositories. It's much harder to find things in institutional repositories.* (Arts & Humanities)
- *People need to be trained how to use them. Once they realise how the benefit far outweighs the effort it will be more successful. Link to research DATA as well as texts.* (Biosciences)
- *Our University lacks the websites that many universities have. The first step before a repository is to organise webpages for all academic staff and research fellows/students. All the academic people and labs should have up to date websites with background/CV of people, published works and ongoing research and optionally make full-text journal papers available online. Currently I have difficulty finding the research trend of professors in our own University. This is a disaster.* (Biosciences)
- *Subject repository not institution repository are the way to go: no one will go through repository institution by institution to find material, it s a useless segregation of output (the publication process is individuals outputting work in a subject, institutions are only the physical address of those individuals).* (Biosciences)
- *Quality control of materials deposited is key.* (Biosciences)
- *Could set up alerts when new material close to one's research interests are uploaded - facilitate cross-disciplinary ventures.* (Science & Engineering)
- *I think long term preservation etc might be particularly important, especially as Research Councils are introducing more guidance/requirements in this area now as part of codes of good research etc.* (Science & Engineering)
- *Providing citation links to articles.* (Science & Engineering)
- *I think it's unlikely that institutional repositories will ever be very useful to me. They might be useful to my university, which indirectly benefits me. But I think they're destined not to play a significant part in my life, and I wouldn't want effort to go into trying to change that.* (Science & Engineering)

- *Subject-based repositories are probably the main source of articles in my field (Physics). While institutional repositories may be in principle helpful for institutions to track publication statistics, I find them to have no scholarly value, at least in my field. It would be a significant job for someone within the institution to keep track of all the high-energy physics papers of which members of the institution are authors and add them to the institutional database, so this is not done, meaning that any publication statistics taken from the institutional repository would be wildly inaccurate. I also note that papers written in my field already require at least four re-formattings, to meet the requirements of the internal database of the experimental collaborations in which we do the research, for the target journal, for the subject database, and for the institutional research group's preprint database if appropriate - this is already a lot of wasted duplication.* (Science & Engineering)
- *I can think of no use for institutional repositories (for my work) which is not already provided in a more global way by arxiv.org.* (Science & Engineering)
- *I never use institutional repositories as prefer to obtain the formatted, corrected version of journal articles directly from the publisher. If my institution doesn't subscribe, then I'd contact the authors directly. Pre-prints are harder to read and authors often don't bother to implement the changes suggested by referees or publishers in the archived article.* (Science & Engineering)
- *(In Edinburgh) we have been exhorted to submit to the repository but the submission process is entirely unclear; the route from preprint to repository, including copyright clearance, should be as easy as possible.* (Social Sciences)
- *I don't agree with the options you gave … I think repositories are a complete waste of time. I can't imagine why any serious scholar would wish to use them. If all institutional repositories ceased to exist tomorrow the loss to scholarship would be negligible.* (Social Sciences)
- *There ought to be resources in place to support the conversion of published papers into a form suitable for use in a repository. I went to a briefing session on the research repository system planned for my own University and was very much put off by the cumbersome nature of the preparatory work which would be involved in contributing to the repository.* (Social Sciences)

Table 19: 'Are there other services you can think of that repositories could develop to broaden their appeal and usefulness to you?' - Comments

Clinical/Health	78%
Social Sciences	74%
Arts & Humanities	68%
Biosciences	68%
Science & Engineering	50%

Table 20: 'I am strongly motivated to publish my research to support the REF, i.e. to place my institution's research at the forefront and to demonstrate its excellence and relevance to society' - Yes

Science & Engineering	39%
Arts & Humanities	35%
Biosciences	32%
Social Sciences	28%
Clinical/Health	24%

Table 21: 'Have you ever accessed an institutional repository in order to find a colleague's research or discover other research relevant to you?' – 'No'

Science & Engineering	21%
Biosciences	17%
Social Sciences	9%
Arts & Humanities	8%
Clinical/Health	2%

Table 22: 'They give the work of the individual researcher more exposure' - 'Disagree'

Science & Engineering	13%
Biosciences	8%
Social Sciences	7%
Arts & Humanities	5%
Clinical/Health	0%

Table 23: 'They broaden the reach of research to others who do not have access to published toll access journals' - 'Disagree'

Biosciences	36%
Science & Engineering	32%
Arts & Humanities	30%
Social Sciences	25%
Clinical/Health	17%

Table 24: 'They reduce the dependence of institutions on increasingly expensive modes of scholarly/research publishing' - 'Disagree'

Science & Engineering	34%
Biosciences	30%
Arts & Humanities	29%
Social Sciences	24%
Clinical/Health	7%

Table 25: 'They represent an exciting new mode of scholarly communication' - 'Disagree'

Biosciences	38%
Arts & Humanities	33%
Social Sciences	33%
Clinical/Health	27%
Science & Engineering	16%

Table 26: 'Institutional repositories will expose more research to plagiarism' - 'Agree'

Arts & Humanities	48%
Biosciences	42%
Social Sciences	38%
Clinical/Health	32%
Science & Engineering	32%

Table 27: 'Institutional repositories may breach the confidentiality of data in some research' - 'Agree'

Response Data: Survey of Subject Librarians from Research Libraries in Nine Countries

Arts & Humanities	21%
Science & Engineering	20%
Social Sciences	13%
Biosciences	11%
Clinical/Health	7%
Average	*14%*

Table 28: 'My role includes a higher proportion of collections management than it used to'

Biosciences	56%
Arts & Humanities	53%
Clinical/Health	50%
Science & Engineering	50%
Social Sciences	47%
Average	*51%*

Table 29: 'My role includes a higher proportion of learning and teaching support than it used to'

Social Sciences	47%
Clinical/Health	43%
Biosciences	33%
Science & Engineering	30%
Arts & Humanities	26%
Average	*36%*

Table 30: 'My role includes a higher proportion of research support than it used to'

- *Increasing interaction with vendors to make online resources more usable*
- *Focused on transition of all three to digital environment*
- *Try to engage faculty more on scholarly communication, but with modest results*
- *e-Science*
- *Varies between subject areas*
- *More focused on going out to the clients' groups rather than receiving them in the library*
- *The emphasis on digital has demonised the face-to-face approach*

Table 31: Other responses to question on ways in which role has changed.

Arts & Humanities	9%
Social Sciences	5%
Biosciences	0%
Clinical/Health	0%
Science & Engineering	0%

Table 32: 'At least a few times a week I provide advice on alternative publishing models, including the institutional repository'

Science & Engineering	13%
Arts & Humanities	6%
Biosciences	5%
Clinical/Health	5%
Social Sciences	5%

Table 33: 'At least a few times a week I promote Open Access'

Social Sciences	13%
Arts & Humanities	9%
Biosciences	5%
Clinical/Health	0%
Science & Engineering	0%

Table 34: 'At least a few times a week I provide advice on compliance requirements for local policies surrounding scholarly communications'

Arts & Humanities	6%
Clinical/Health	5%
Social Sciences	5%
Biosciences	0%
Science & Engineering	0%
Science & Engineering	0%

Table 35: 'At least a few times a week I design and implement improvements in institutional repository functionality'

Biosciences	25%
Arts & Humanities	16%
Clinical/Health	14%
Social Sciences	14%
Science & Engineering	0%

Table 36: 'At least a few times a week I participate in the procurement and development of new tools and services for research'

Biosciences	13%
Arts & Humanities	6%
Social Sciences	5%
Clinical/Health	0%
Science & Engineering	0%

Table 37: 'At least a few times a week I provide advice on the management of grey literature'

Arts & Humanities	6%
Social Sciences	5%
Biosciences	0%
Clinical/Health	0%
Science & Engineering	0%

Table 38: 'At least a few times a week I provide advice on data management'

Social Sciences	10%
Arts & Humanities	6%
Biosciences	0%
Clinical/Health	0%
Science & Engineering	0%

Table 39: 'At least a few times a week I provide advice on alternatives to traditional peer review'

Social Sciences	24%
Arts & Humanities	13%
Biosciences	13%
Science & Engineering	13%
Clinical/Health	5%

Table 40: 'At least a few times a week I provide advice on copyright and intellectual property rights'

Social Sciences	24%
Arts & Humanities	13%
Biosciences	13%
Science & Engineering	13%
Clinical/Health	5%

Table 41: 'At least a few times a week I provide assistance in 'marketing' research through alternative means, such as web 2.0'

Biosciences	13%
Arts & Humanities	6%
Clinical/Health	5%
Social Sciences	5%
Science & Engineering	0%

Table 42: 'At least a few times a week I provide bibliometric and usage statistics reporting – for both traditional and alternative publication models'

- *Print journal subscriptions*
- *Reference ('students and interns (cheap labor) are being used')*
- *Resources replaced by links*
- *New book information (in paper form)*
- *Reading lists (done by another Library department)*
- *AV*

Table 43: 'Which traditional subject librarian services have been superseded by new ones in your library?' - Comments

Yes	43%
No	57%

Table 44: 'Is the demand for new services recognized by your institution with changes made to support these new arrangements, including changes to your job description?'

- *'Reallocation of more traditional skills (staffing enquiry points, cataloguing and classification) to Senior Library Assistant level and below'*
- *'Repository work'*

Table 45: 'What changes have been made?' - Comments

Yes	84%
No	16%

Table 46: 'Do you think that such services should be part of your job description?'

- *Academic Liaison Team*
- *Repository Management in Systems*
- *Dedicated staff (for Repository and Research Support)*
- *Digital Assets Team*
- *Programmers/Developers outwith the Library*
- *Research Development/Support (with Library involvement)*
- *Research Enterprise Division*
- *College research support services*

Table 47: 'If 'no', say who should provide them'

Yes	67%
No	8%
Don't know	24%

Table 48: 'Do you believe web 2.0 tools and approaches are important in supporting research dissemination?'

Yes	47%
No	53%

Table 49: 'Do you personally make use of web 2.0 tools to support research dissemination?'

Blogging	20%
Social networks (including Twitter and Facebook)	19%
Wikis	13%
Social Bookmarking	10%
Newsgroups and forums	9%
Instant Messaging	9%
Online video	7%
Podcasting	6%
Photo-sharing	5%
Wikipedia	1%
Bibliographic software social tools	1%
Virtual worlds	1%

Table 50: 'Which web 2.0 services do you use to communicate and disseminate research?'

Yes	55%
No	15%
Don't know	30%

Table 51: 'Do you think the development of subject-focused social networks (online communities) is an important role for liaison librarians for research dissemination?'

- *Set up and administer an Allied Health Researchers online community*
- *Instigated a national email group of tertiary liaison librarians in pharmacy*
- *Around 60-70 subject-focused mailing lists developed to disseminate research/funding news over the past 10 years or so*
- *Run blogs as part of particular courses, and offer blogging and wiki workshops which are embedded in assessment*

Table 52: 'Can you give examples of social networking initiatives you have taken?'

Positive	84%
Negative	16%

Table 53: 'Would you say that the use of web 2.0 tools and social networks in support of your role has been a positive or a negative experience?'

Appendix D: ERIS Project NLS Work-package 3 activity

1. Scottish Research Aggregation

Version	Change	Author(s)	Date
0.1	Initial draft	James Toon	15/06/2010

Purpose

This document outlines the broad scope of the NLS development contribution for the creation of a Scottish research data aggregation service as part of work package 3 of the ERIS (Enhancing Repository Infrastructure in Scotland) project for approval. The document is not intended as a technical or functional specification.

Mission

(a) To establish an operational service that collates and aggregates together the metadata and full text contents (where available) of Scotland's institutional repositories (IR), making the aggregation available as a queriable electronic resource for subsequent discovery by the research community and general public.

(b) To normalise metadata acquired from HEI repositories in order to provide feeds from the aggregation database that can be queried by third party systems for the development of broader national or international systems, without loosing the integrity or discoverability of the original object.

(c) To provide a prototype service that feeds the aggregation as content into the NLS long term preservation archive.

Background

The universities of Scotland have invested heavily in the development and implementation of repositories over the last 5-10 years to serve both as a record of institutional research output, but also as method of increasing the availability and access to knowledge via open access.

There is however no collective record of Scottish institutional output that can be used for identifying local research capability and to help facilitate mutually beneficial relationships between Scotland's research institutions and the Scottish business community.

Much of the research contained in these repositories is only available via the local systems that support them, and with uncertain prospects for many institutional repositories in the current financial climate, the need to ensure long term preservation of repository content has never been more important.

Aims and Objectives

The core aims of the project are;

- To provide a service, hosted at NLS, that aggregates and collates the available contents of Scottish HEI institutional repositories via available data collection points such as OAI-PMH
- To provide a set of administrative mechanisms that allow for the scheduling, configuration, error logging auditing and other such functions that allow for serviceable operation
- To develop an ingest process which will inspect and normalise the metadata acquired from across Scottish HEI via the aggregation in such a way as to ensure that the integrity of the original resource is not compromised

- To normalise the data in such a way as to make it available as a CERIF[12] data source (nb: resultpublication[13] entity) – restricted to journal article type only at present.

- To make the full normalised aggregated repository of content available as an OAI-PMH data source for other external third party harvesting.

- To index all content and metadata and develop a search interface that allows for users to search the contents of the aggregation in a flexible way, either via simple full text search, or more complex faceted searching. The search should, at a minimum, allow for filtering of discovery by institution.

- To integrate into the search, as available, the HILT[14] subject discovery web services being developed by the CDLR (Centre for Digital Library Research) at Strathclyde University.

- To retain as much branding as possible during the discovery process for the originating institution.

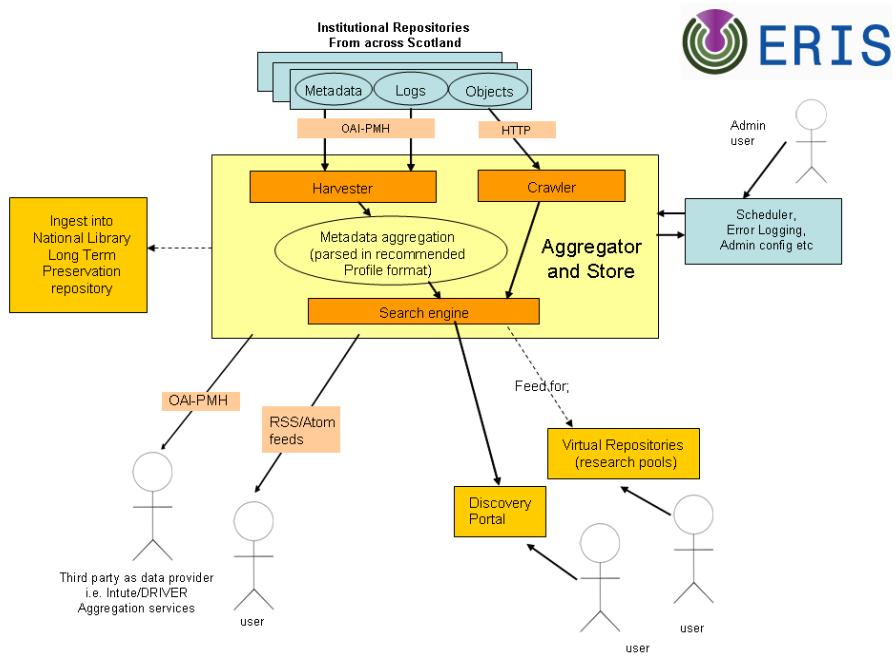
- To ensure that on discovery, the user is redirected to the host repository original data source 'jump page' contained at the local repository

- To provide this aggregation of content as an electronic resource for use with other NLS discovery systems that would be of benefit to the Scottish research community (i.e. links to business via Scotbis)

- To provide a demonstrable feed of aggregated content into the NLS long term preservation repository ingest mechanism.

[12] CERIF – Common European Research Information Format http://www.eurocris.org/cerif/introduction/
[13] http://www.dfki.de/~brigitte/CERIF/CERIF2008_1.1FDM/Physical/MImageB.html#ERD_1
[14] HILT High Level Thesaurus project http://hilt.cdlr.strath.ac.uk/

Figure 1. Concept diagram of aggregation



Measures of Success

The following will be considered as objective measures of success for the project

- A working and production ready online system that undertakes and manages the aggregation of metadata and full text content from Scottish HEI that provide access to their repository contents
- A set of administrative functions that allow for customised scheduling, error logging and configuration of harvesting parameters, plus an audit log of CRUD activity.
- A customer facing discovery UI that allows for queriable searching, but not least to be able to search and filter results by originating institution (see New Zealand[15] and Irish national research[16] portals as examples)
- Data contained in the repository is normalised as a core internal schema, to be defined by the project. [NB: probably use CERIF format as spinal scheme, but requires discussion]
- The availability of the data contained in the aggregation repository as an OAI-PMH data source for use by third party services
- The availability of data contained in the aggregation repository in CERIF compatible format via a suitable API for use with virtual repository solutions for research pools
- The retention of institutional branding for each search result, both in the results lists and in more detailed information pages.

[15] http://nzresearch.org.nz/index.php/index
[16] http://www.rian.ie/

- The link from the result must redirect the user to the institutional jump page for the item contained in the local institutional repository from which it originated.

Testing

The project has limited resource available for direct testing on the development work and so the development lead must pay particular attention to unit testing and in developing appropriate test harnesses for activity and ensuring that code is regularly reviewed

The project will be required to create a number of use case scenarios that will be used to develop a testing plan for the project. This testing will provide evidence required to sign off against the measures of success defined. The primary testing strategy is to ensure development of objective measures at every opportunity.

Assumptions, Dependencies and Constraints

Assumptions

The project assumes the existence of suitable harvesting technology (either as developed in house via IRIScotland work, or via third party solutions)

Dependencies

The project is dependent on the input from the Front of House and ERS teams in defining the operational workflow for reader (member) authorisation

Successful testing of the systems is dependent on their being a full test environment in place and available both internally and externally to NLS

Constraints

The project is subject to the following constraints;

Resource constraints

The development work at NLS is constrained to the available development resource, funded up to £36349 (Vat Inclusive) via the ERIS Project.

Time constraints

The development project can run for no longer than to the end March 2011 (ERIS project completion date), but ideally will need to be complete by end December 2010 to allow for project evaluation and closure work.

Non- financial resource constraints

The project will be subject to the following non-financial constraints;

- Software applications used in the development of the system architecture will need to support the Library's preferred development and vendor solutions

- Software produced should be made available as open source according to JISC project funding requirements; (http://www.jisc.ac.uk/fundingopportunities/opensourcepolicy.aspx)

Legal/Policy Constraints

- NLS must observe local institutional access policies for digital materials harvested
- NLS should prepare a suitable statement of liability for contents

Project Interfaces

ERIS Work package 1 activity

The ERIS project is undertaking a user study into the motivations for researchers and research pools to use institutional repositories. The study will provide information which may affect the scope of development activity.

ERIS Work package 2 activity

ERIS Work package 2 is undertaking a feasibility study into the provision of a contracted long term preservation service, to be hosted by the National Library of Scotland. The outcomes of this study may indicate demand is sufficient to raise the long term preservation prototype work in this development to a production service.

ERIS Work package 3 activity

The project will need to interface with the development activity underway at CDLR in the development of a HILT web service to improve resource discovery.

The project also wishes to undertake a number of other projects, based on the investigations from the WP1 study. This will be subject to available resource, and may or may not include the NLS development resource.

CRISPool project (St Andrews - http://www.crispool.org/)

The ERIS project is participating in a partnership with another JISC funded project called CRISPool, which is also looking in to the development of research information management services for research pooling. The ERIS Project is producing CERIF based data sources at the local level for a number of institutions. This development takes the opportunity to prepare an alternative route for institutions to comply with CERIF formatting without having to make local changes to their repository systems.

Project Governance

The workpackage is being run as part of the ERIS Project, and is subject to governance requirements as laid down by the project board.

Roles and responsibilities:

- ERIS Project Manager: James Toon (UoE)
- Workpackage 3 Manager: James Toon (UoE)

- Development lead: Richard Claydon (NLS)
- System Developer, TBA (NLS)
- System testing, TBA (NLS)

Communications

Richard Claydon (RC) will be the project team member responsible for internal and external communications as follows;

- Fortnightly checkpoint report. RC to produce and distribute to project team as above
- Fortnightly project review (checkpoint) meeting (JT, RC plus others as required)
- A Monthly highlight report to be submitted to Workpackage 3 members (JT)
- Work package 3 delivery group meetings as dictated by the ERIS project

Project Controls

The project will be managed according to broad PRINCE2 principles, but predominantly through management of scope and via risk and issue management.

Costs

Costs and resources

The project is being undertaken by a dedicated resource, paid for by the ERIS Project, but managed by the National Library of Scotland

Schedule

The exact details of the schedule are yet to be finalised, but the broad time constraints laid down by the Programme are as follows;

	Description	Start Date	Finish Date
1.	Work package development	1st July 2010	31st March 2011
2.			

These dates are high level estimates and will be confirmed during the initiation process

Appendix E: Work Package 3 - Technological Enhancements

Product Descriptions

25 August 2009

Version	Change	Author(s)	Date
0.1	Initial draft	James Toon	25/8/2009

Mission:

To provide a clear understanding of the products that are required for delivery as part of work package 3 (Technological Enhancements)

Aims/Objectives:

To set expectations for the project and the delivery team members when working on the delivery of work package 3. The product descriptions provide key information about what is to be delivered, in what format and how the products will be assessed for sign off.

Product Descriptions:

ID 1	**Title:** User scenarios and use cases
Purpose: The use cases describe the overall scope of delivery and scenarios for the work package	
Schedule: Draft completion of Scenarios by end August 2009, Full sign off of cases and scenarios by end Sept 2009	
Compositions: For each of the 5 identified enhancements there will be; • User Scenarios (General description and user scenarios) • Actors involved • Use case workflows • Use cases	
Derivation: The user scenarios and use cases are derived from carrying out desk research into the areas of development	
Format and Presentation; The user scenarios will be in the form of a word document	
Allocated to: WP3 team	
Quality Criteria/method: The user scenarios will be signed off by the WP3 delivery team, having been reviewed and accepted as complete and realistic descriptions of the technical enhancements that are proposed.	

ID 2	**Title:** Use case realisations/feasibility for each module in Dspace and/or Eprints
Purpose: The use case realisations represent the high level designs and approaches that will enable the development teams to construct and demonstrate working prototypes of the functionality proposed	
Schedule: Completion and demonstration of functionality and supporting design documentation in place by end march 2010	
Compositions: For each of the 5 identified enhancements there will be; • User case realisations (i.e. technical translations of the proposed use cases) • System designs	

• Working system prototypes • Testing scenarios and test cases • Supporting technical documentation
Derivation: The use case realisations and prototype developments are derived from the user scenarios
Format and Presentation; The use case realisations will be in the form of word documents (or similar) and the feasibility prototypes will be developed code, compiled in a demonstrable manner (i.e. prototype versions of the systems proposed that can run and demonstrate expected results)
Allocated to: WP3 team (technical)
Quality Criteria/method: the quality criteria for the use case realisations will base on their ability to translate the use cases and scenarios into high level system designs. The developed system prototypes will need to meet expectations set out in the test cases and provide an objective demonstration of system functionality against the user scenarios

ID 3	**Title:** Implemented (developed and tested) enhancements for Dspace/Eprints as per requirements
Purpose: To take the prototypes developed in the previous phase, and, having reviewed them against user expectation, refine the implementations into serviceable and workable products that meet the user's needs.	
Schedule: Completion and demonstration of functionality and supporting design documentation in place by end September 2010	
Compositions: • Prototype user assessments • System design refinements • Final system test cases refined from previous phase • Development of 'production ready' code • Delivery of code as open source project (as per JISC requirements) • Release documentation	
Derivation: The implemented enhancements are derived from the prototyping phase, and also based on the results of the WP1 – user engagement.	
Format and Presentation; The use case realisations will be in the form of word documents (or similar) and the feasibility prototypes will be developed code, compiled in a demonstrable manner (i.e. prototype versions of the systems proposed that can run and demonstrate expected results)	
Allocated to: WP3 team (technical) and ERIS project management	
Quality Criteria/method: The developed prototypes will be made subject to gateway review by users. The refined and 'productionised' code will be signed off against the refined test cases The finalised products will be released as open source, and made available as a wider project. The project will request advice from the JISC Open Source Software Advisory service on the sign off of these packages	

ID 4	**Title:** Validation report against DCC methodology for designing and evaluating curation and preservation experiments
Purpose: The DCC methodology will be used to validate the satisfactory transfer of objects between repositories following the implementation of the OAI-ORE transfer.	
Schedule: Completion of review against DCC methodology complete by end December 2010	
Compositions: • Methodology test cases (use cases) • Defined test experiment • Test results (from PLANETS test bed environment) • Evaluation report • Supporting documentation	
Derivation: The product requires completion of the OAI-ORE component of WP3	
Format and Presentation; The use cases and evaluation report will be in the form of word documents	
Allocated to: WP3 team (technical) and the Digital Curation Centre	
Quality Criteria/method: The product will need to demonstrate that the experiment followed the method described, and the evaluation report will be peer reviewed and signed off on a subjective basis.	

ID 5	**Title:** Validation report against user identified requirements in line with WP1 (WP3 summative evaluation)
Purpose: To undertake a review of the developed functionality with the repository user community, in particular where the needs have been developed to meet the researcher users.	
Schedule: Completion of review against WP1 findings by end December 2010	
Compositions: • Peer review of developed functionality • Review with user stakeholders (researchers/managers/administrators) • Review data • Evaluation report (summative)	
Derivation: The report will be derived from work in WP1 (User engagement) and from the implementation of the technical enhancements in WP3	
Format and Presentation; The report will be in the form of word document	
Allocated to: WP3 team	
Quality Criteria/method: The validation report is an internal Summative evaluation of the developments against the projects objectives. The report will be subject to peer review and will be subjective in nature	

ID 6	**Title:** A report of follow on actions for WP3 activity, post project completion
Purpose: To report on the conclusion of work package 3, and to make recommendations as to how the outcomes of this work package should be maintained on a practical and manageable basis.	
Schedule: Completion during Jan-Mar 11 project closure period – Due March 2011	
Compositions: • Follow on actions report and recommendations • Stakeholder communications (presentations and/or articles)	
Derivation: The follow on actions report is derived from all other WP3 activity, but in particular from the DCC and project evaluation work	
Format and Presentation; The follow on actions report will be a word document and prepared presentations (powerpoint or prezi)	
Allocated to: WP3 team (technical)	
Quality Criteria/method: The follow on actions report must produce a realistic set of recommendations and actions that have objective measures of success. The report will be subject to peer review and project board review.	

Appendix F: UWS/SAC Basic Brief

Purpose: to provide a low cost and lightweight solution to enable institutions to have access and use of a repository.

Its low cost because it's a shared setup, and they have no real ability to do anything other than simple interaction.

The idea is that the users of this shared repository can collaborate together in order to plan and develop functionality. They develop a community approach to managing a repository. Sharing development costs, sharing advocacy materials and approaches and so on.

Models for operation are the White Rose repository http://eprints.whiterose.ac.uk/
NITLE http://dspace.nitle.org/

Basic List of Requirements

To set up a shared repository service – a community model in which a single repository serves x number of institutional needs.

The basic set up consists of a single Dspace instance, with some minimal branding. Version 1.6.2 assumed as default (and as installed – V1.7 also available and may be worth upgrading to in the event that the service is retained)

The initial repository uses the default deposit workflow offered by the software (out of the box)

Simple collection structures negotiated which focus on the deposit type rather than collections based on organisational structures

The central admin role is performed by SDLC (or equivalent)

Each repository instance has a small number of users (one or two) and will be subject to mediated deposit in the first instance.

Community/collection roles are specific to each institution, so one cannot edit or deposit into another.

Need to set up a distribution license as part of the process – using something default, but should be agreed by both parties

Check the roles for the submissions process. For example, users in community X should not be able to see submissions in community Y

Need to set up handle server for communities.

Run google sitemap code against each repository for indexing (although have to wait to ensure that each repository user is happy to have their material made 'live')

Install remote log file viewer function as developed by Paul so that errors can be observed and managed.

Questions

Can we ensure that each jump page is branded to the institutional owner of the collection

Can we improve the details of the metadata record – current short record format seems 'too short'

Do we need to have separate handle servers for each community?

Wish lists (general ideas if we have time/resources)

Would be good to set up authority control for authors, based on a local namespace and using HR record from UWS/SAC

Set up Easydeposit client with Sword2 which can be deployed easily at each institution for academic deposit rather than having to get them to log into the repository each time

Set up google analytics for repository instances, and test to see if we can provide analytics access to each community rather than on each whole repository

Add publication lists functionality (is it possible to use the code developed in the PR for this? Is it portable?)

Discussions required with UWS/SAC

Talk to institutions around usage of metadata fields. Need to set standards at the outset.

How can we get collection contents out for re-import into a.n.other repository? Export via OAI-ORE ?? What are our options

Do we need to make any substantial changes to workflow?

Collection structures

Appendix G - ERIS Project - NLS Work-package 3 activity/CDLR Development Scope (Eprints User views)

Version	Change	Author(s)	Date
0.1	Initial draft	James Toon	05/08/2010
0.2	MePrints comments	William Nixon	24/08/2010
0.3	Revisions following JT/WN review	James Toon	26/08/2010

Purpose

This document outlines the broad scope of the development for the development of a package of user facing enhancements in Eprints as part of work package 3 of the ERIS (Enhancing Repository Infrastructure in Scotland). This document is not intended as a technical or functional specification.

Mission

(a) To establish an enhanced user facing view of repository contents in Eprints using the MePrints plug-in[17] developed under the JISC rapid development stream, as a base for improving the user experience and encouraging use of the repository as a tool in the research life cycle. This would be done in both an ERIS Demonstrator EPrints service (hosted by the NLS?) and at the University of Glasgow.

(b) Based on work carried out at the University of Glasgow in Enlighten and via the Enrich project, create MePrints widgets to show associated projects and funder for authors

(c) To provide and enhanced email notification system that allows for the sending of an email on a pre-configured basis (such as once per week/month/quarter) containing a digest of information about a users repository items, including

- Statistics about the views and downloads of an item – by period and cumulative to date
- List of items that are held as metadata only as a reminder to upload full text
- Method of quickly adding files to items

Background

The universities of Scotland have invested heavily in the development and implementation of repositories over the last 5-10 years to serve both as a record of institutional research output, but also as method of increasing the availability and access to knowledge via open access.

However, it is widely acknowledged that despite the best efforts of repository and library staff, the academic community has not become fully engaged with the use of repositories, either as means for increasing availability of research or as a source of information in the research life cycle.

The ERIS project has worked with user communities in an attempt to establish what it is that is preventing the widespread take-up of repositories as a research tool, and a common reason has been established as 'a lack of personal control' in the management of their content. This work package is aiming to develop a number of enhancements based around the Eprints MePrints plug in, which was developed to ***'enrich the repository users experience by giving them more interesting and informative home within the repository environment?'***

Aims and Objectives

The core aims of this work package are;

- To implement the MePrints plug-in for Eprints[18] as (at least) a proof of concept at the University of Glasgow and in an ERIS EPrints Demonstrator

- To create a default MePrints homepage, which will enable users to view their content via a range of MePrints widgets.

[17] http://allaboutme.eprints.org/
[18] http://wiki.eprints.org/w/MePrintsOverview

- To create a public profile page for registered users which will show publications such as their most viewed and most downloaded as well as their College/School affiliation
- To create a MePrints widgets which will show projects and funders for authors
- To develop an enhanced email notification system which is configurable by the user to provide a digest of repository content information on a regular scheduled basis
- As a subset of the email notification function, and for each user, collate usage statistics at the item level from within Eprints and present for inclusion in the digest email.
- As a subset of the email notification function, and for each user, collate and present a list of item records that are held in the repository as a metadata only record as a reminder to upload full text, and include this information optionally in the same digest email, or as a separate email.
- As a subset of the email notification system, provide a quick upload feature that allows users to click on the reminder items in the email and in the MePrints interface, that will allow for the quick upload of an item against and existing metadata record.

Measures of Success

The following will be considered as objective measures of success for the project

- A commitment to implement the solution by project partners following development and testing
- A working installation of Eprints with MePrints populated with staff and research data
- A working customisable and user 'home page' with a suite of MePrints widgets to replace the default "Manage Deposits" screen
- A public profile page showing publications as well as departmental information e.g. Department, School and e-mail address
- User receives a scheduled, automated email from Eprints containing simple usage stats for their repository items
- User receives in the same email, a list of items, distinguishing between those that are full text and those that are metadata only
- The user is able to update the metadata only records to include a full text document

Delivery approach

Work on the package will commence from beginning of October.

The core MePrints Widgets to be installed include:

- User Details - Shows the username, biography, expertise and other core data fields.
- User Picture - Allows the user to upload an image of themselves for their profile.
- User Actions - Adds a box containing all of the possible user actions.
- Recently Published Items - Shows items that the user has recently published in the repository.

- Item Issues - Shows any issues with items, owned by the user, that have been discovered by an issues audit.
- Quick Upload - Allows the user to upload a file and attach it to a new record in their inbox, ready for publication on the repository later.
- Most Popular Items - Looks at the number of times a users items have been accessed and displays a rundown of the most popular.
- IRStats Mini Dashboard - If the repository has IRStats installed this widget can display data from it and allow the user to look at a number of different statistics regarding their publications.

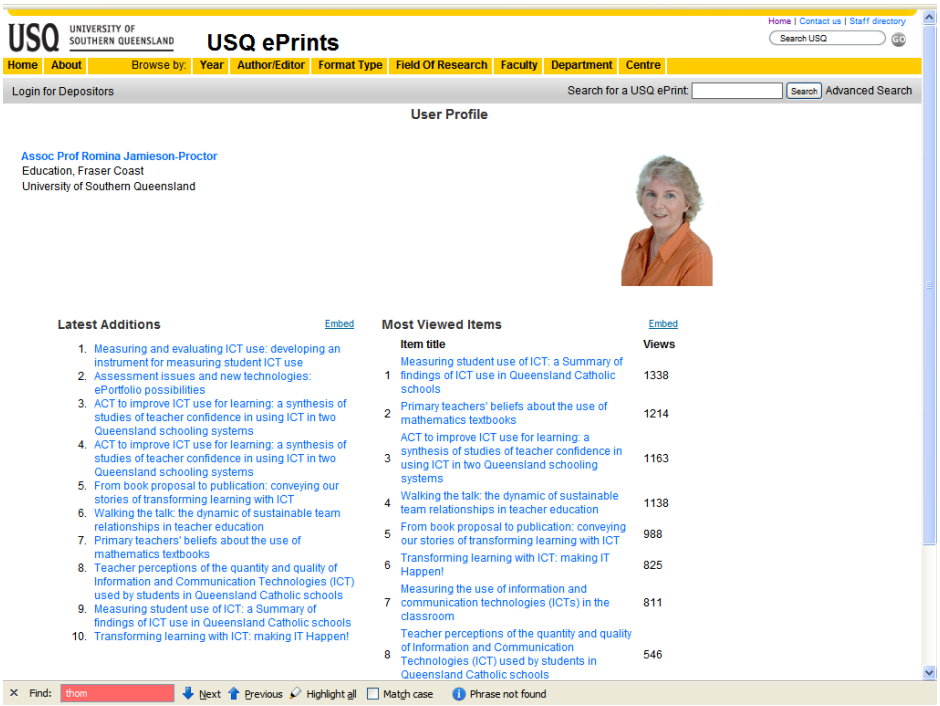
The University of Glasgow (and QMU) is using IRStats and would be very interested in the practical use of the IRStats widget..

Phase 2 – MePrints widget development

In addition to these the development of widgets to show Funder and Project information would be very useful – and could draw on data gathered by the University of Glasgow as part of the Enrich Project. The University of Glasgow has data on a wide range of projects and funders including the Medical Research Council, the Wellcome Trust and the Engineering & Physical Sciences Research Council.

Phase 3 – Design of default profile page

An example page from the from the University of Southern Queensland.



By default MePrints shows papers deposited by a user, but it is possible to modify this with code provided by the EPrints team which changes the definiton of ownership in eprints to authorship so

that a paper can be belong to multiple authors. The code was written by Tim Miles Board and is available on the MePrints Blog[19].

This issue of ownership and the author to publication links has been identified by both the University of Glasgow and Queen Margaret University.

Phase 4 - Development of Email notification facility

Develop an enhanced email notification system that allows for the sending of an email on a pre-configured basis (such as once per week/month/quarter) containing a digest of information about a users repository items.

Phase 5 – User feedback

Demonstrate features and functions to potential Scottish collaborators (such as QMU and Napier) and engage with academic users to gain feedback to service improvements.

Testing

The project has limited resource available for direct testing on the development work and so the development lead must pay particular attention to unit testing and in developing appropriate test harnesses for activity and ensuring that code is regularly reviewed.

Repository staff at the University of Glasgow can demonstrate the work at the University's Research Systems Group as well as to key College and School staff who they are are working closely with.

The project will be required to create a number of use case scenarios that can be used to develop a testing plan for the project. This testing will provide evidence required to sign off against the measures of success defined.

User scenarios can focus on a number of the key widgets for instance the use of the quick upload feature and evidence of engagement as a result of the ready availability of download and view data for individual staff.

The project should also ensure that users are provided with the opportunity to comment on the developments and approaches throughout the course of the work package. The team will endeavor to recruit and engage on an informal basis with academic users based out of the University of Glasgow and Strathclyde, and also seek input from other Eprints users in Scotland – notably at Queen Margaret University and Napier University.

Assumptions, Dependencies and Constraints

Assumptions

The project will use Eprints 3.2 as the repository software with MePrints 1.3[20] add-in

CDLR will set up and maintain appropriate source control mechanisms as required by the project, and to comply with JISC project policy[21] requirements

[19] http://blogs.ecs.soton.ac.uk/MePrints/2009/11/25/MePrints-out-with-the-users/
[20] http://files.eprints.org/501/
[21] http://www.jisc.ac.uk/fundingopportunities/opensourcepolicy.aspx

The project assumes that MePrints is the most appropriate vehicle for delivery of enhancements proposed.

Dependencies

Successful testing of the systems is dependent on there being a full test environment in place and available both internally and externally to NLS. The University of Glasgow also has a development installation of EPrints which can be used for testing.

Constraints

The project is subject to the following constraints;

Resource constraints

The development work at CDLR is constrained to the available development resource, funded via the ERIS Project.

Time constraints

The development project can run for no longer than to the end March 2011 (ERIS project completion date), but ideally will need to be complete by end December 2010 to allow for project evaluation and closure work.

Non- financial resource constraints

The project will be subject to the following non-financial constraints;

- Software applications used in the development of the system architecture will need to support the CDLR preferred development and vendor solutions
- Software produced should be made available as open source according to JISC project funding requirements; (http://www.jisc.ac.uk/fundingopportunities/opensourcepolicy.aspx)

Scope Exclusions

Non-academic Eprints repositories are considered out of scope for this workpackage (i.e. National Museums)

The project is focusing on repositories of research publications. Learning and teaching repositories are out of scope.

Interfaces

ERIS Work package 1 activity

The ERIS project is undertaking a user study into the motivations for researchers and research pools to use institutional repositories. The study is not anticipated to have any impact on the development of the aggregation functions proposed in this project.

Enrich Project – University of Glasgow.

This ERIS project work package is participating in a partnership with another JISC funded project based out of the University of Glasgow (Enrich),

Governance

The workpackage is being run as part of the ERIS Project, and is subject to governance requirements as laid down by the project board.

Roles and responsibilities:

- ERIS Project Manager: James Toon (UoE)
- Workpackage 3 Manager: James Toon (UoE)
- System Developer, Stephane Sechaud (CDLR)
- Project officer (Assistance with business analysis and Enrich outcomes), Willam Nixon (UoG)

Communications

Stephane Sechaud (SS) and James Toon (JT) will be responsible for maintaining internal and external communications as follows;

- Fortnightly checkpoint report. SS to produce and distribute progress update to project team as above
- Fortnightly work package review meeting. Probably via telephone but F2F if possible (JT, SS plus others as required)
- A Monthly highlight report on progress to be submitted to Workpackage 3 members (JT)
- Work package 3 delivery group meetings as dictated by the ERIS project

Project Controls

The project will be managed according to the ERIS Project governance (which are based on broad PRINCE2 principles), but predominantly through regular communications and management of scope and via risk and issue management.

The development work will be managed using AGILE approaches deployed by the CDLR.

Costs

Costs and resources

The project is being undertaken by a dedicated resource, based out of CDLR, and paid for by the ERIS Project.

Schedule

The exact details of the schedule are yet to be finalised, but the broad time constraints laid down by the Programme are as follows;

	Description	Start Date	Finish Date
3.	Work package scoping	1st Oct 2010	31st March 2011* TBA
4.	Estimations and phase requirements development		
5.	Code Development		
6.	System and user testing		
7.	Project closure and follow on implementation actions		

These dates are high level estimates and will be confirmed during the initiation process

Risks

The following risks have been identified at project commencement.

Key risks faced and mitigation strategy for addressing them			
Risk	**Impact**	**Likelihood**	**Mitigation**
Ambiguity of scope	High	Medium	Ensure adequate dialogue and sign off at project start up
Impact of work package 1 study (motivation of researchers in repository use) may raise discussion over scope of aggregation services	High	Low	Stick to rigorous issue/risk management following agreement of initial scope and enforce chance control process if required.
Unavailability of Technical staff/support	High	Low	Ongoing documentation of implementation

Appendix H: Enhancing Repository Infrastructure in Scotland - Work Package 3 – Development Report (University of Strathclyde)

Phase 1 – Use Cases & Scenarios

The task of writing up use cases and scenarios was split between Stephane Sechaud and Anu Joseph. The process involved speaking with the repository administrators and end users to gain valuable feedback on issues that would improve both the functionality and usability of using the institutional repository by identifying regularly submitted requests and complaints by users to repository managers.

It was identified that a regular request was for modifications to the subject metadata available in the repository. Most EPrints repositories have a subset of the Library of Congress Classification Outline available at install to be imported as the subject metadata scheme. Amending this subject metadata usually involved editing a flat file containing the hierarchal data of the subject vocabulary. This issue was adding to the workload of repository administrators due to the time and effort involved in configuring and maintaining a local subject vocabulary. We proposed that extraction of subject metadata from a central remote source would both remove the workload of repository administrators maintaining the local subject metadata, while also providing a greater selection of headings or terms from an authoritative provider of subject metadata.

It was also identified that end users wanted a greater level of accessibility to records in the repository by using metadata filters to aid the search and retrieval process. To demonstrate this we choose the keywords metadata field as a source of user generated metadata that would be useful to expose to users in a more tangible way i.e. providing keywords as auto-suggestions at the point of deposit, and generating new browse views against keywords. We proposed that by changing the default keywords metadata field format to have multiple values, instead of a single concatenated string, would be able to provide keyword auto-suggestions at the point of deposit. This would also allow us to expose the keywords metadata field as another browse view to access the repository content.

Other partners within the ERIS project produced a scoping brief that described the features of both an email reminder system for a) scheduled emails with statistical information about users' deposits in the repository, and b) providing an opportunity to remind users to upload full-text records for their metadata only deposits. It was also expressed that it would be beneficial to have this statistical information available for exposure on both the users' homepage and public profile pages provided by the MePrints plugin, available for the EPrints repository software.

Phase 2 – Development

The development process mainly involved researching the EPrints repository software and its API via the documentation provided online via the EPrints website [1] and wiki [2]. After discussions it was decided to prototype and test demonstrators for solutions to the issues identified during processes detailed above in the Phase 1 section.

Subject Metadata

To tackle the subject metadata issue, it was decided that linked data architecture would be more suitable to provide subject metadata to repositories as it would negate any extra management on the repository administration team by offloading the maintenance of the subject metadata vocabulary back to the authoritative sources of the subject metadata vocabularies. The linked data architecture would ensure that subject metadata available to users for attaching to their records would always be up to date and comprehensive.

To implement this it was necessary to research and select an appropriate software package to provide the triple store functionality necessary to implement a locally working demonstrator of how the proposed solution would work. This involved the testing of multiple software packages to find one that was both suitable and manageable for development staff to use for the implementation. Once the

triple store had been installed and configured, it was then necessary to obtain the subject metadata vocabulary data in a linked data format and import this data into the triple store.
Due to IPR and licencing issues, Library of Congress Subject Headings (LCSH) was selected as it readily available in an appropriate format. The Dewey Decimal Classification (DDC), which is considered to be one of the most desirable subject vocabularies due to its decimal notation which provides for greater scope for being able to inference and group results, it is still only partially available in a linked data format due to licencing issues. However it has been stated that both the DDC and Universal Decimal Classification (UDC) schemes will be available online in a linked data format in the near future.
Once the triple store and dataset was configured and ready for use, it was then a matter of developing scripts that would utilise the EPrints architecture to provide new subject metadata fields in the repository and allow for these fields to be exposed at the point of deposit, with dynamic auto-suggestion of the subject metadata. These fields also required to be made accessible to users as browse views.

Keywords

In tackling the issue of keywords, we identified that due to the variability of how the keywords field was handed in repository installation that it would be problematic to have the enhancements proposed automatically created as this would involve changes being necessary on the repository side and it was felt that this was better left to repository administrators to decide how they would like to move forward with implementing this solution. A basic PHP script was created as an example of how one could take the existing single value keywords field, delimited by commas, split the values, perform some de-duplication and normalisation of the values and populate and new table for keywords to be used for auto-suggestion at the point of deposit. This was deemed an inappropriate solution as it would increase the workload of repository administrators, requiring them to maintain the keywords data to be offered as auto-suggestions due to the field being handled as a single value, and would not allow for an appropriate browse view to be generated.
It was then decided that a more useful solution would be to change the keywords metadata field to contain multiple values instead of one. This would allow for a much easier implementation that would not require any maintenance on the part of the repository administrator to maintain the keywords metadata for auto-suggestions, while also allowing for a browse view to be generated for this field.

Email Reminders

A scoping brief was produced by William Nixon and James Toon that defined the specifications and functional requirements for the implementation of an email reminder system that would enable repository users to receive regular email notifications with links to upload full-text attached to their records without full-text available. Statistical information was also specified as a feature that would provide useful information to users that would detail the usage of their deposits in the repository.
After reading the scoping brief document it was decided to redefine the specification to remove the stipulation that the email reminders settings form be displayed as a MePrints widget. This change was made to ensure the email reminders feature would not depend on MePrints being installed on the EPrints repository as this was an unnecessary requirement. The proposed solution was to produce an EPrints plugin that would provide the form to configure the email reminder settings by the user, a bin script in Perl that would provide the heavy lifting of control the scheduling, generation and sending of the email reminders. This bin script would then be added as a scheduled task on the server to run once every month. All text content to be used in the email was separated into an XML phrase file, as described by the EPrints development guidelines, to ensure support for localisation.

MePrints Widgets

The development of the MePrints widgets mainly involved reading the existing documentation that comes with MePrints and inspecting the existing widget scripts to determine how they worked. Once it was understood how they were implemented we then developed the two widgets to display the funders and projects with the most items deposited for that user. The statistics where gathered via SQL queries to the database table maintained by the EPrints software to log usage statistics for the repository.

Software Used

All development work was carried out using free or open source software. Servers were installed as virtual machines using Oracle VirtualBox [3], with the exception of the triple store which required to be hosted on a physical machine due to performance requirements when querying the triple store. The triple store software used was Arc2 produced by Semsol [4]. The Ubuntu Server Linux distribution [5] was used as the operating system for the development server virtual machine.

Issues & Lesson Learned

It was felt that the changes made to EPrints to enable the keywords field to hold multiple values, required further discussion and that agreement would be needed among repository administrators to take this forward. It was also felt that, if this development was considered a positive one, a script would be useful for repository administrators to use that would enable them to convert their existing single value keywords field into a multiple value field within EPrints without losing any data. This would be a relatively simple task, however time has not been allocated to this, but it is expected that it will be sourced soon from the EPrints community.
During the course of our development work we relied heavily on the documentation that exists online for the EPrints repository software and it's API. Although there is documentation available, we often found that due to either issues relating to the configuration of our repository, or the variables in our development server meant that we were unable at first to implement our solutions using all the of EPrints API calls that we thought necessary. It was later discovered that these issues originated from a bug found in the XML::LibXML Perl library that is a requirement for the EPrints repository software. This meant we had to change code to have it behave in the way we expected it to. These kinds of issues are to be expected, but due to the fact that the EPrints API is still being actively developed and that the documentation isn't comprehensive and complete, it took us longer to complete some of our work that we would have hoped. Thankfully the EPrints tech list provided invaluable help in responding to our email enquires about the problems we experienced during the development phase.
In regards to the external subject metadata solution there is a concern regarding the availability of SPARQL endpoints that provide the linked data with which we provide the subject metadata to the repository. This is a requirement that cannot be negated as it is intrinsic to the architecture of the semantic web and linked open data. Initially the SPARQL endpoint for DDC was unavailable and we were forced to undertake the task of implementing our own triple store/SPARQL endpoint server to let us demonstrate the functionality. Luckily the DDC SPARQL endpoint is currently available, but there exist the potential for this to disappear in the future, so we recommend that an institution or official body should step forward and volunteer to host SPARQL endpoints that are available for HE/FE institutions to use in the future development of services which can tap into the useful and growing popularity of sematic web technologies and linked data.

References

[1] – http://www.eprints.org/
[2] – http://wiki.eprints.org/w/Main_Page
[3] – http://www.virtualbox.org/
[4] – https://github.com/semsol/arc2/wiki
[5] – http://www.ubuntu.com/business/server/overview

Appendix I: ERIS work package 2: case studies

Contents

- Purpose and intended outcomes 1
- Curation policy at the University of Aberdeen 1
- Curation policy at the University of Glasgow 4
- Curation policy at the University of St Andrews 7
- Summation 9

Purpose and intended outcomes

The purpose and intended outcomes from these case studies comprise four key elements:

i. To explore, document and compare the current and planned policy frameworks for digital preservation and in particular data curation within a sample group of Scottish universities.
ii. To test the appropriateness of the ERIS Curation Policy Framework (CPF) in a 'live' situation and subsequently modify it to reflect the feedback received.
iii. To provide feedback to institutional stakeholders that can be used in the adaptation of the CPF to meet local needs.
iv. A report to the ERIS Project/JISC that will describe the process, issues, observations and recommendations of these case studies.

1. Curation policy at the University of Aberdeen

Methodology

The primary contact, the institutional repository manager/cataloguing manager at the University Library, identified interviewees from amongst the principal stakeholder groups[22], suppliers of hosted repository services and informed/influential researchers at the University of Aberdeen. In addition to the primary contact, the interview group consisted of a Policy Adviser and Senior Policy Adviser from the Policy, Planning and Governance office, the Freedom of Information Officer, the University's Records Manager, the Head of Graduate School in the College of Physical Sciences and the Digital Development Manager from the Scottish Digital Libraries Consortium (SDLC).

A meeting was held with four interviewees in Aberdeen, including the primary contact. Subsequent telephone conversations were held with a further two, followed by a meeting with the SDLC representative.

The University of Aberdeen repositories

The institutional repository, *AURA*, is based upon DSpace software and is maintained under a hosting arrangement with SDLC. It contains full-text research output (currently articles and working papers); there has been recent pressure from academic staff wishing to upload databases and musical scores/compositions.

In collaboration with the University of St Andrews, Aberdeen has also introduced a Current Research Information System (CRIS), which is based upon the proprietary *Pure* software. It holds no full-text research output but with a focus on management information it includes metadata describing research publications, research activities and impact records, together with administrative data drawn from the institutional Human Resources and Grants & Contracts databases.

A second repository, *Digitool*, produced by the library systems company Ex Libris, is home to digitised historic collections, some born-digital collections, e-theses, and past examination papers. The institutional *Virtual Learning Environment* (VLE), managed by the Centre for Learning &

[22] Those who would be responsible for introducing and formalising a policy, those responsible for implementing it and those with responsibility for adjacent policies and procedures such as data protection and freedom of information.

Teaching, also contains a wide range of learning objects. The VLE employs Learning Object Metadata standards, whilst Digitool uses Dublin Core and MARC21 (although both are expressed and made available for harvesting as Dublin Core).

Established in 2004, *AURA* was initially popular with only two departments and deposit remained patchy. With the introduction of *Digitool* in early 2008 it was planned to migrate *AURA* content across. This plan was revoked with the ascendancy of the *Pure* project, which absorbed the existing RAE-driven publications database in 2009. In April 2010, *AURA* was moved to SDLC hosting and *Pure* went live. The two systems operated independently for a few months until a link was established in July 2010, the crosswalk between *Pure* and DSpace being based on the Metadata Object Description Schema (MODS). It is now only possible to ingest content into *AURA* via *Pure*'s submission mechanism as part of a single workflow (it cannot be done independently of *Pure*), with the two systems having become inextricably linked.

RAE 2008 had been a crucial driver in this sequence of events. An earlier aim had been to join up the repository and the publications database via a CRIS but since *Pure* had the capacity to link to a DSpace repository the University took the opportunity to upgrade and outsource *AURA* to SDLC, making the connection between publications output and research business processes and creating an opportunity for the demonstration of research impact. Interestingly, whilst a CPF would focus on academic output, the *Pure* CRIS could tend to blur the boundaries with administrative data, which would have to be addressed in any policy framework tailored to such an installation.

In terms of uptake, *AURA* currently captures around 20% of the institution's full-text output, and *Pure* 80-90% of metadata only. Typically, arts and humanities research is under-represented. In consideration of long term sustainability, the University of Aberdeen is committed to maintaining a repository. With its fundamental strategic goal to become one of the world's top 100 universities, repositories are seen as key in terms of raising the institution's research profile.

University of Aberdeen publications policy

Aberdeen's Principal, a social statistician with a Research Council background, is a proponent of Open Access. With a publications deposit mandate on the Senatus agenda for academic year 2010/11, library staff are preparing for its imminent introduction. Guidance on "Good practice for research" is already available to academic staff and the University Records Manager is leading on the development of guidance for the management of datasets, in connection with which some data security roadshows have been organised.

Data curation at the University of Aberdeen

Most activity in this area has been concerned with textual research outputs but recent high-profile news stories (e.g. 'Climategate') have made senior management aware of data management as an issue deserving attention. The Library and the Policy, Planning and Governance unit are now working to develop procedures and systems to support a data management ethos. In this, the need to maintain access to research data after a researcher has left the University is a key motivating factor.

Policy developments

It is apparent that the development of data policy will only be taken forward when a business case with clear and demonstrable benefits to the University can be made. According to the Senior Policy Adviser from the Policy, Planning and Governance unit, the University would be keen to introduce an institutional data policy as a means of driving efficiencies, which would be managed strategically by the Centre. However, a complex and diffuse investigatory and decision-making environment for resolving knowledge and information issues provides a particular challenge.

For example: the *Pure* project is driven by the Vice Principal for Research and Innovation. It is overseen by a Research Information System project board comprising the Vice Principal, the Heads of Research from each of three constituent Colleges (Arts & Social Sciences, Life Sciences & Medicine, and Physical Sciences), the University Librarian, and the Director of IT. The Vice Principal for Research and Innovation is responsible to the University Court, the institution's governing body. Papers have also been presented to the Advisory Group on Information Strategy (AGIS), which advises another Vice Principal having responsibility for information strategy. The project team that reports to the *Pure* project board includes representatives from the IT Directorate's Applications Management Division (who are concerned only with Pure and not AURA)

as well as Documentation and Helpdesk teams, a project manager (also from the IT Directorate), and members of the Policy, Planning and Governance unit.

Policies relating to the institutional repositories are managed by the University's Policy, Planning and Governance unit. Representatives of this unit were interviewed to establish the background and future plans for the repositories as well as for data management.

The Records Manager advises on policy and practice development across the institution but is principally concerned with the application of Records Management in support to administrative functions. This has recently included advocating a digital preservation strategy, with the goal of attaching named owners to all records (digital and analogue), together with disposal dates built upon retention schedules. The Records Manager's function includes the development of policies for managing records held by individual sections of Administration (e.g. Human Resources, Finance or Registry) and to promote these policies within the University. Advice is also given on issues including suitable storage, filing methods, file naming, folder structure, version control, data security, retention, preservation and disposal. Breen will also be looking at standards and procedures for research records in the future.

The Data Protection and Freedom of Information Officer has not been aware of the University having received any identifiable Freedom of Information (FOI) requests for research data but accepts that this is likely to change. Just as FOI requests to other universities for data about learning and teaching have been driven primarily by controversy, it is anticipated that a similar situation will pertain to requests for research data. The case of the Queen's University Irish Tree Rings was cited as an example, and there is a constant expectation that requests for information about animal research at the University will be received.

Beyond the central Policy, Planning and Governance unit, the perception in the Library is of a somewhat fragmented policy landscape, where policies are developed to meet specific events or needs without much evidence of joined-up thinking behind them. Whilst it is recognised that support for research needs to be undertaken by a range of support services on any campus, and that close co-ordination and integration are the key features of any successful approach, the research support infrastructure is considered to be somewhat deficient. There is, for example, no dedicated research support librarian and College Research Directors have instead to liaise with subject librarians working to a broader portfolio.

Gaining approval and achieving buy-in to data curation at the University of Aberdeen will require an evangelical promotion of the benefits from having a CPF. In this, the University's e-Research Strategy Group may have a key role to play. This working group has been delegated by the Committee for Research, Income-Generation and Commercialisation to submit its recommendations in March 2011, which will ultimately be reported to the University Court, where the presence of e-research as a vital element of the current University strategy will be emphasised. Nonetheless, interviewees for this case study expressed the view that any policy proposal would be more likely to be accepted if it is perceived as coming from outwith the University, particularly where researchers can observe other large Scottish universities undertaking similar measures.

Data management infrastructure at Aberdeen

The provision of central services in support to an institutional data management policy would be difficult to achieve as currently there is no data curation expertise in the Library or the IT Directorate. Any potentially relevant expertise is scattered throughout the University and can be hard to identify. The impetus for introducing systematic data management services has also been threatened by a recent programme of voluntary severance, which has hit non-academic staff numbers the hardest, thereby reducing the cohort most likely to have a key role in policy development and the provision of supporting infrastructure.

In addition to this deficit in the human infrastructure there are concerns about the technology infrastructure, with respect to the ready provision of appropriate storage capacity. This is an interesting dilemma, as we are more used to hearing that storage costs are not a significant barrier. The challenge is compounded by the diversity of attitudes and mindsets with respect to the management of data, which are at different stages of evolution according to researchers' subject domains.

2. Curation policy at the University of Glasgow

Methodology

An investigation into the curation policy for scholarly research output at the University of Glasgow comprised structured interviews with the Assistant Director for Research & Learning Support Services (19-10-10) and the Director of the University Archives (16-11-10). They were subsequently amplified by a conversation with the Vice-Principal for Research and Enterprise on 04-02-11.

The University of Glasgow repository

The centrally supported university repository, *Enlighten* (http://eprints.gla.ac.uk/), was launched in April 2006 as an outcome of work undertaken by the *DAEDALUS* project, a three year library-sponsored initiative funded by the JISC to explore the composition for a model institutional repository. It has an *EPrints* software platform and is able to support full text and open access. All content including metadata in *Enlighten* is open to public view (subject to specific publisher embargoes). The institutional repository operating procedures do not contain any measures for the disposal of material deposited

In 2008 the University's senior management mandated deposit to it of scholarly publications and University of Glasgow theses. Self deposit into *Enlighten* is sometimes undertaken by academic staff; otherwise, deposit is either by departmental administrators or materials are sent to the library for them to effect ingest.

The repository is administered by dedicated staff within the Library as a core library function. It is regarded as a fundamental institutional service and is staffed by a Repository Manager with cataloguing staff who work exclusively on repository records. The Repository Manager reports to the Assistant Director Research & Learning Support Services, who in turn reports to the University Librarian. The Assistant Director also works closely with the Vice-Principal for Research and Enterprise, who chairs the Research and Planning Strategy Committee and provides the main link to both the research agenda and senior management.

University of Glasgow publications policy

In response to the introduction of the Research Excellence Framework (REF) as a replacement for the Research Assessment Exercise (RAE), particularly in consideration of the REF's inclusion of an element of bibliometrics, the University recognised the value of research publications as an important asset by introducing a publications policy (http://www.lib.gla.ac.uk/enlighten/publicationspolicy/) with the aim of

- increasing the visibility of research publications produced by staff employed by or associated with the University of Glasgow;
- ensuring that research outputs are prepared and curated in a way that helps maximise their value in terms of the extent to which their external use can be measured bibliometrically.

This policy was approved by Senate on 5th June 2008, at which point *Enlighten* was recognised as the institutional repository and platform for the University's publications database, into which staff are expected to deposit a copy of peer-reviewed, published journal articles and conference proceedings (subject to publisher conditions). Since then the deposit of full text publications has been 'patchy' but there is near-100% compliance in the provision of metadata about publications. The Library is in the process of developing procedures and protocols for each of the four Colleges to cover workable measures for collection and deposit.

With maintenance of an institutional publications database confirmed as a core strategic aim, the need for sustainability in the face of constrained and potentially diminishing resources has led to the decision that other library activities would be deprioritised in order to preserve and continue the necessary level of resource.

Data curation at the University of Glasgow

The institutional publications policy does not address the deposit, management or curation of research data and at an institutional level strategies for the management of data have been slow to reach the senior management agenda, which has been preoccupied with a major organisational restructuring of the University.

A Digital Preservation Advisory Board chaired by the Vice-Principal for Research and Enterprise (who also chairs the Research and Planning Strategy Committee), which has a brief that extends across teaching, research and institutional records produced in digital form, had previously created a working group with membership from the relevant information support services (Library, IT, etc.). The group's task was to examine the issues arising from the preservation of electronic records and research data, including some investigation of individual research projects. Recommendations were produced but it was concluded that the time was not yet right for taking them to senior management. However, in February 2011 the Research and Planning Strategy Committee agreed to reprise this scoping study, having resolved that much of what it had covered has since been taken forward in a collaboration with the University of Cambridge, through the auspices of the JISC-funded Incremental Project. Perhaps most significantly, the Incremental team at Glasgow had developed an online portfolio of *Data management support for researchers* (http://www.gla.ac.uk/services/datamanagement/), a series of web pages designed to help research staff plan, create, organise and look after their data, including advice on where best to seek assistance.

When challenged, the University of Glasgow's research community claims that it *knows what to do* with the final output of data from a research programme but there is little evidence that effective curation principles are being applied during the research lifecycle. At the same time researchers complain that they cannot find advice and assistance on data management, although it is available from several sources, which points to the need for a serious communication and education programme. The work started by the Incremental team is therefore likely to be taken seriously.

Researchers have been known to seek assistance from the university's Archives when at the end of a project they wish to have their data stored, but Archives do not have the necessary resource to meet these needs. Consequently, large volumes of electronic data are held in the Schools, according to varying degrees of maintenance and security and with little or no régime of disposal in place. The practice of individuals keeping everything is viewed as some kind of insurance against disaster, unpredictable need or changes in circumstance, yet no method to ensure the long-term usability of retained data is applied.

There is no dedicated central facility for the preservation of scientific data and, generally, research data is held locally within the sub-units of the four Colleges (although there are no declared data repositories maintained in the Colleges). Whilst it is known that the issues surrounding the effective management of data are recognised within the Colleges, not least in the sometimes controversial context of data ownership, information services staff acknowledge that there is little likelihood they can be tackled successfully at an institutional level until there is complete or greater buy-in to the mandate for the deposit of publications.

One other very basic factor contributing to slow progress in the development of a data service is that neither the library nor IT services can currently offer specific expertise in data management or curation. A further complication is the perceived necessity for discipline expertise in the selection and appraisal of data for inclusion in a data repository; this would have to be undertaken at a College level in order to ensure that appropriate discipline knowledge is brought to bear, a potentially unwelcome consideration in a climate of cost constraint and reducing staff numbers.

Across the institution one can define three distinct areas where electronic data is being gathered and used: teaching, research and institutional records. Each of these has its own *champion* through whom it may be feasible to address each area separately, thereby reducing the scale of the problem and with an opportunity to develop best practice for export to the other two.

Policy developments

A further JISC-funded project, *ENRICH* (2009-2010), has aimed to improve the integration of the *Enlighten* repository service with the University of Glasgow's research management system (http://www.jisc.ac.uk/whatwedo/programmes/inf11/sue2/enrich). The *ENRICH* programme includes an ambition to

- develop clear policies and workflows with academic departments and Faculties to ensure a sustainable and ongoing flow of content into the institutional repository;
- deliver an agreed institutional preservation policy, which will be formulated in discussion with IRIScotland (the emergent pan-Scotland repository infrastructure) and the National Library of Scotland.

The information policy environment therefore has some precedence: a mandate covering scholarly publications, a practical response to the demands of the REF and the commitment of the Research and Planning Strategy Committee to consider salient issues. Further policy development would be routed through the Research and Planning Strategy Committee to the Senior Management Group and thence to Senate, as was the case with the publications policy.

The development of data policy at Glasgow would not be initiated without the expression of serious interest or need by senior university management. Drafting would most likely commence as a discussion between Library and IT managers. Whilst the relevant services have already begun to engage with the academic community on matters concerning data curation, as witnessed by creation of the *Data management support for researchers* portal, the ownership of policy has still to be clarified, although it would be issued under the authority of the Vice-Principal for Research and Enterprise and/or the Principal.

At an operational level, the introduction of an effective data management framework will require the combined effort of the Archive, Library and IT teams, but currently there is incomplete buy-in to data management as a high priority issue amongst the members of these groups. A robust and authoritative policy alone would therefore be unlikely to succeed unless it is coupled with a programme of advocacy and education, whilst the policy itself must offer practical guidance and solutions.

A key factor influencing the pace at which policy is being developed (or not) is the apparent reluctance of the research funders to monitor and police compliance with agreed data management plans. Despite a majority of the major funders having introduced specific requirements for the submission of data management plans as part of a research grant proposal, as yet they are putting no pressure on grant holders to prove their compliance once plans are approved and money disbursed. Should this situation change, the focus on arrangements for managing research data at the University of Glasgow, as elsewhere, will be greater.

Data management infrastructure at Glasgow

The University of Glasgow is currently active in the introduction of new corporate IT systems (HR, student records, etc.) but the focus is on functionality and technology rather than maintenance of the data itself. Much of this is due to the lack of capacity to resource the kind of higher level activity represented by digital curation. So, in terms of current developments in data management, one will find plans for storage but not for preservation, access and re-use. Consequently, what and how much data will survive longer-term is an unknown.

Institutional records are increasingly born digital but whilst the institutional archive has established mechanisms for the selection and appraisal of records, no resource has been provided to enable long-term storage and use. Capacity, in terms of the availability of both specialised staff and storage facilities, has not been matched to any perceived needs for applying organised data management. With funding having become particularly constrained as a result of changed economic circumstances, increased investment is unlikely unless it can be demonstrated that improved data management will contribute to the alleviation of pressures on institutional resourcing.

3. Curation policy at the University of St Andrews

Methodology

Discussion and analysis was undertaken with a range of stakeholders identified in conjunction with the Deputy Director of Library Services, focusing on individuals who would be responsible for introducing and formalising curation policy, those responsible for implementing it, those with responsibility for adjacent policies and procedures, suppliers of hosted repository services and informed and influential research staff.

Three of the stakeholders were interviewed at St Andrews; a further two were subsequently consulted by email and a meeting took place in Edinburgh with a representative of the Scottish Digital Libraries Consortium (SDLC). In all, six people were consulted for this study.

The University of St Andrews repositories

The institutional repository is *Research @ St Andrews: Full Text*, which holds the full-text of e-theses and some voluntarily-deposited research papers. The repository is based upon DSpace software under a hosting arrangement with SDLC. *Research @ St Andrews: Full Text* is linked to the Current Research Information System (CRIS), *Pure*, which was procured and implemented in partnership with the University of Aberdeen (Case Study 1). Each institution now has its own separate installation of *Pure*.

Additionally, a pilot Fedora-based *Arts Data Archive* is being developed and, whilst not normally referred to as a data repository, a Moodle virtual learning environment (VLE) **(**used primarily by the English, Geography, Physics and Psychology disciplines) contains material used in teaching.

Between 2002 and 2005 St Andrews took part in the HaIRST project, a consortium of three Scottish universities and ten Glasgow FE colleges, which investigated the design, implementation and deployment of a pilot service for institutional resources. An EPrints repository system was introduced at St Andrews but by 2005 the burden of providing support locally became untenable and it was decided to join the Scottish Digital Libraries Consortium (SDLC). The repository software was changed from EPrints to DSpace and the service outsourced to SDLC. In 2007 the St Andrews repository, now known as *Research @ St Andrews: Full Text*, was linked to a Research Expertise Database used to support the 2008 Research Assessment Exercise (RAE). The Research Expertise Database has now been succeeded by the *Pure* CRIS system, with a configuration similar to that found at Aberdeen (i.e. a *Pure* + DSpace combination hosted by SDLC).

Postgraduate researchers can deposit their theses directly to the DSpace repository, or Library staff will do this on their behalf. The deposit of theses has been mandated since 2006. Researchers may also self-deposit publications, although this is done via Pure. The main Library catalogue also holds bibliographic MARC records of the nascent *Arts Data Archive*, which constitutes the beginnings of a data asset register.

The *Pure* information system (the public view onto which is called *Research @ St Andrews*) holds project administrative information, records of research activities and CVs, publications metadata, and some DOI links to full-texts held elsewhere (e.g. publishers' sites, OA sites, PubMed etc). It also stores the full-text of publications ingested via *Pure* and accessed from their home in the DSpace repository. The interoperability of *Pure* and DSpace is based on the Metadata Object Description Schema (MODS.

Resource has been made available to the Library to carry out repository advocacy and support, which has led to an increase in deposits. Repository activities enjoy cross-unit support from the Library, the Research Policy Office, and the Business Improvements division. However, the view was expressed that if resources became markedly tighter the CRIS system supporting Research Assessment would enjoy higher institutional priority than the full text repository, despite the latter experiencing growing support from the institution as a whole.

University of St Andrews publications policy

Despite the existence of established repository services, no mandate is currently in place with arrangements for data.

Data curation at the University of St Andrews

A limited number of developments in managing electronic data have been driven by IT Services (ITS). Through their Arts Computing Advisor (a post not replicated for other disciplines) they already support Arts Faculty researchers in the preparation of data management plans (DMPs) and they are in the process of developing a dedicated, Fedora-based electronic archive for Arts research data. Principally, this archive will contain data that would previously have been deposited with the now defunct Arts and Humanities Data Service (AHDS).

The dual purpose of the *Arts Data Archive* is to demonstrate that St Andrews is meeting funder requirements and to protect and preserve fragile, unique electronic resources. The archive will accept software code, research data (starting with XML/MySQL/Postgres databases and digital images) and documentation. It will link to metadata in *Pure* rather than duplicate content.

In terms of sustainability, the new *Arts Data Archive* has attracted positive interest from the Chief Information Officer (CIO), who has been promoting it to the Vice Principal for Research. This

interest has yet to produce a firm commitment to further development funding. Restructuring of the IT Services and Business Improvements units is underway and may result in the merger of academic and administrative computing, which in turn could have an impact on the scope of the new *Archive* by broadening its potential user community. Notwithstanding this prediction, the *Archive* will be rolled out during the summer of 2011, its first content being the outputs from the *Records of the Parliament of Scotland* project.

The provision of data management services that cover research and administrative data was perceived as an attractive option by the interviewees for this study, but with much yet to be achieved with the publications repository and with the information focus for the next two years being upon the REF, early progress in the introduction of new protocols and services for research data is unlikely. That said, it was recognised that if the submission of data were more strongly embedded within the REF it would present an important influence. It was also recognised that an unexpected loss of important data would also be likely to provide some impetus to the introduction of principles for data management and curation.

Policy developments

St Andrews has a powerfully autonomous, devolved culture with few institution-wide policies, none of which address digital preservation. Two related policies have been drafted but not yet ratified: an Information Security Policy and an Administrative Data Policy. Additionally, new policies for the *Arts Data Archive* are being developed that are based on former AHDS policies, but the scope of these may expand if the new *Archive's* coverage expands to include non-Arts subjects.

The absence of a mandate for the deposit of publications is unlikely to change soon, since resistance to such a move is anticipated from the vocal community of Arts researchers, who hold particularly strong views about the ownership of research publications.

Any new policies relating to information and data would go via the CIO to the Principal's Office for approval, although the process by which drafts make the journey to adopted policies is ultimately dependent upon the nature of the policy in question, as well as the enthusiasm and approach taken by its 'champions' in senior management. In the case of a Curation Policy Framework the likely champions would be the CIO and the Deputy/Vice Principal for Research but whilst the introduction of a digital curation/preservation policy is supported by the Library, there is little current support from senior management. On a practical level, the successful roll-out of such a policy would require the allocation of funding and human resources and with financial support being voted on a strategic basis the initiative would be dependent on being driven from the top down rather than bottom up. In the absence of essential Vice Principal leverage, St Andrews is unlikely to be introducing such a policy in the near future.

Data management infrastructure at St Andrews

Administrative data is handled by the Business Improvements (BI) division of central administration. A full time Data Architect, also acting as project manager for the deployment of *Pure* in BI, has a global remit covering all administrative data, including information relating to research projects, activities and outputs.

Pure is linked with Tribal Technology's *SITS* student record system, which includes data about undergraduate and postgraduate students. There is a pervasive sense of integration, easily achieved, perhaps, as a consequence of St Andrews' small institutional size. *Pure* is treated as the 'golden' source of research output bibliographic data and research activity data; it also integrates with the DSpace full-text repository. The University's Research Policy Office and the Library liaise with HR, Registry and the Research Funding Office as data providers.

Within the institution, understanding of key digital preservation frameworks such as OAIS is felt to be virtually non-existent; hence tailored training courses would be essential. Neither has St Andrews shown itself to be sufficiently strong on applying good practice routines such as data backup and, although IT Services could be expected to address this situation, there is a shortage of available effort to undertake effective advocacy. The provision of training in good data management is further frustrated by the internally devolved nature of the institution, where training needs are determined by individual heads of department.

This study found that the expertise needed to implement a curation policy is deemed to be in place within the University but new specialist skills may be needed for actually developing and supporting

data management. But as St Andrews already outsources some repository functions a similar approach may be considered for data.

4. Summation

The principal aim of these case studies was to explore, document and compare the current and planned policy frameworks for digital preservation and in particular data curation within a sample group of Scottish universities, whilst at the same time testing the appropriateness of the ERIS Curation Policy Framework in a live environment. Without exception, even though there are degrees of variation, little evidence was found of co-ordinated data curation (or even serious attention to it) at a corporate level, with a complete absence of data management or curation policy.

The prevalent focus on documents rather than data still exercises the providers of repository operations in each of the three institutions that were studied and, even there, notwithstanding the introduction of a mandate for deposit at two of them, the rate of deposit ranges from patchy to restrained. This is not untypical but suggests that the effort involved in continuing to develop an effective document or text repository could detract from extending activity into the realm of data.

That said, two of the study sites have invested significantly in research information systems and these may well prove to be the seed from which a strategic interest in managing data will grow. As new opportunities to exploit the functionality of these systems will inevitably present themselves, with the potential for the Research Excellence Framework to evolve into the capture and measurement of data and data citations and with increasing competition between institutions to demonstrate research value, it is at least possible that the merits of nurturing trusted and re-usable research data will be recognised and an appropriate infrastructure put in place.

At Aberdeen it was noted that the Library and the Policy, Planning and Governance unit are actively working to develop procedures and systems to support a data management ethos; at Glasgow the Library-based Incremental project has enjoyed success in drawing the attention of researchers to the more critical needs of data custodianship, which it continues to support; whilst Information Services staff at St Andrews have made a significant thrust into the data management arena with their *Arts Data Archive*.

Each of these initiatives has emerged out of concern by professional information or data-orientated staff, none of whom occupy positions of influence at a senior management level. But at that level, interest has at least been aroused in each of the three institutions studied, with policy groups at two of them already examining which steps might now be taken.

We observed that the direction of travel towards a managed data infrastructure has been in response to external financial pressures, whether in order to meet funder requirements at St Andrews or, in the case of Aberdeen and Glasgow, in strategic preparation for the Research Excellence Framework. It is likely that the new and significant investment in human and technological infrastructure that is implied by a data curation policy will, similarly, need to be legitimised by those same issues of compliance and competition, with its momentum governed by the equation of strategic necessity.

More than one interviewee referred to the 'opportunity' that a catastrophic loss of data would bring to the cause of systematic data management. This should not be judged as mere cynicism. Recent examples at other UK universities bear witness to the acceleration of engagement with data curation that follows a major fire or breach of data security.

With regard to the ERIS Curation Policy Framework, it was not possible to perform an actual test in an environment where curation has not become an established process. However, interviewees were enthusiastic in endorsing much of its approach and contributed further comment on issues of practicality that have been incorporated into the final draft.

Whether interest in data curation evolves slowly and of its own accord, or in response to the exigencies of government and other agencies, or as a direct response to a disaster, the groundwork by the professionals that is described in these case studies will at least mean that whilst there is currently no curation framework in place, knowledge of the aims and principles for curation has

already begun to coalesce. In that case, the ERIS Curation Policy Framework may at last be tested in earnest.

Appendix J: ERIS Quality Plan (Autumn 2010). WP 4.

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
WP1 – Enhancing Researchers' Engagement with Repositories							
Established ERIS Researcher and Research Pool communities	Developing relationships/links between various stakeholder groups, i.e repository managers, research pools, liaison libs, SCURL heads	Ongoing thorough project	Evidence of communications and positive relationships?? What can demonstrate a good rel with researchers? Ongoing communications? Evidence of ongoing participation - attendance at events etc. Programme of stakeholder events for each group across the two years of the project?	Positive evidence of community aspects, i.e tacit knowledge transfer, anecdotal accounts of the benefits of bringing the groups together for practical purposes.	Set up and convening of INSTREPMANSCOT group for repository managers, set up of Liaison Librarians group, Setting up of IRIScotland council, ongoing community development via networking in Scottish HEI and amongst research pools. *** Requires plan for community development – as part of final recommendations to SCURL**	Anecdotal via summative eval? Evidence of communications to stakeholders	Ongoing . Some concerns over ability to sustain community work within the scope of the existing project.
Project Website to be used for ERIS user engagement and collaboration	Communication of project success, and store for relevant project documentation and outputs		Inclusion of analytics into site to demonstrate reasonable visit rate (and via network location)	Install analytics plug in for word press required to track usage (if seen as required)	Require metrics and evidence of readership.	Website **(P)**	Completed (but requires constant attention) **[UPDATE OCT 2010 – SITE TAKEN DOWN DUE**

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
							TO SERVER VIRUS. NEEDS REBUILD]
Researcher engagement survey method and questions(institutional)	Survey to be directed to liaison librarians via SCURL for redistribution to researchers, with the expectation of 10 responses back from each institution - total 300 expected responses.		Response rate and sufficient spread of subject disciplines across Scottish HE. Also full coverage		Section in report detailing method and questions	Agreements between work package managers, discussed via Basecamp site.	Completed
Researcher engagement survey method and questions(pools)	Via mixture of formal/informal meetings with research pool directors/administrators	Ongoing across project	Some survey coverage as a direct consequence of main WP1 survey. Mainly through direct contact with research pool directors and administrators.		Meetings held with all participating pools on a number of occasions. Have attended and joined the PAN meeting group, made up of pool administrators. Workshops in plan to discuss research data and other issues with academic staff (in plan)	Meeting notes held in project folders for use in WP1 reporting. Now written up for WP1 reporting. Needs conversion to blog posting	Ongoing
Compiled and ordered focus group data and simple analysis	Organise research pool and research group focus sessions and/or telephone interviews, with questioning based on RIM work via OCLC	In Feb. Use outputs from initial use case scenarios established via WP3?	Curated data - stored in datashare Analysis plan – expected outcomes or key points expected		Commencing from 10th March.	Audio files - Published datasets (via Datashare) **(P)**	Focus groups completed. D

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
~~Interim report of survey results~~	~~At end feb? need to base around ability to run focus group sessions effectively~~	~~Likely to need to combine interim and final reporting due to time constraints. Two risks to be raised, around the inability to review and re-action survey, and also the impact the survey will have on WP3~~ ~~*technically late – was due for end June, but now likely to be made available end August 2010~~	-	~~Blog posts out by end August to support final survey report~~	~~Blog posts on ERIS site, plus posts on Enlighten (Glasgow) blog re focus groups and OCLC Hanging together blog for survey work~~	~~Produce as blog post series – used to 'test' report with user community~~	**ISSUE No. 19 – Accepted as sufficiently low risk**
Completed Researcher event (to be held at Scottish Agricultural College)	Large focus group session, based on SAC researchers and distributions	Completed in Jan 2010	? positive contribution from members of the meeting?		Notes from meeting - written up and being fed into project reporting	Meeting	Completed
Continuing community engagement plan (strategy to maintain momentum and communications post survey)	In addition to project communications plan, start to distribute regular updates on project progress via internal systems at institutions and via liaison/libs and SCURL	Across project	Pass formal responsibility over to SCURL committee (or propose that they let the ERIS project and IRIScotland council groups convene?)		Development of communities (such as repository managers/liaison librarians groups, and evidence of events/meetings	Report section for inclusion into WP4 business planning **(P)**	Ongoing – summary plan will be delivered as part of final WP1 project report, and elaborated on in final project report
A Final report on the WP1 survey activity and community engagement		~~End march 10~~ Now due for completion at end August 10	Part of publications list for project		Final report – in 'publishable' format. Need to raise this question with the	Published report at end of project (supporting	Not yet

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
					team. Have suggestion to use DCC Style guide ***Report should be made available in project repository**	WP4 planning) **(P)**	
Summative review reports of ~~completed~~ proposed WP3 deliverables against user needs.	Report on whether or not the wp3 deliverables meet the needs of the users, and what in fact should be the development priorities to support user needs	End September 2010			Project board agree and sign off report	Paper for project board as part of formative review. Publication in repository on approval	Needs inclusion in WP1 reporting. JT to discuss with WP1 leaders.
A report of follow on actions for WP1 activity, post project completion		March 11	Part of publications list for project		Will also form part of the appendix of final report and as part of WP4 business planning activity	Report section **(P)**	Not yet started
WP2 – Enhancing Curation and Preservation Processes within Institutions							
An agreed method and approach to surveying Scottish HEI's curation policies	Done - survey has been developed and in testing, preparation for early Oct release.	**End Sept 09**			Forms part of the final report	Report section **(P)**	Completed
Compiled and ordered survey results/data		**End Dec09**	curate and deposit in datashare		Need to discuss with	Published datasets (Datashare)	Completed
Analysis of results		**End Feb 10**				Report section **(P)**	Completed

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
Report on survey of Scottish HEI curation policies		**End March 10**	Part of publications list for project		Distribution of copies, putting report up on project website	Published report **(P)**	Completed
Recommended approach for development of curation policy framework for Scottish Institutions		**End April 2010**			Internal agreement at DCC	Report section	Completed
The Developed Draft Policy Framework		**End April 2010**	Part of publications list for project			Project toolset	Completed
Set of policy framework user scenarios and approach to testing the success of implementation	Testing of policy framework against real	**End March 2010**	Based on actual test subjects. Need to establish who would be willing to take part in this via SCURL		Use case reports published (conf presentations??)	Blog posts and case study reports for inclusion in final framework version	Staffing issues at DCC (new start came and went)
A final agreed policy framework for recommendation by the ERIS project		**End (uncertain – following on from case studies)**	Revised version of policy framework document based on implemented use cases		Distribution of bound copies to participating HEI? Or via Stakeholder groups as previous reporting	Revised version of project output if required	
~~Requirements documentation and Use cases for ERIS metadata profiles for digital objects~~	-	~~**End jan10**~~	~~Need to question the ambitions of these deliverables. Will require quite a lot of time in DCC/CDLR~~	-	-	-	-
~~Design of object metadata profiles to meet requirements (technical and administrative metadata)~~	-	~~**April 10**~~	-	-	-	-	-
~~Test scripts to prove applicability of metadata to selected use case~~	-	~~**July 10**~~	-	-	-	-	-

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
~~scenarios~~							
Report of recommendations for curation policy~~ and metadata approaches for implementation.~~		**July 10**	Essentially this is the WP2 'final report'		Published recommendations	Deliver as article/blog series?	**ACTION JT:** Discuss how we will disseminate WP2 activity in general.
Report of recommendation for long term preservation of IR contents and data, run out of NLS	Feasibility assessment of long term repository as back-up store	**Oct 10**	Can only be theoretical at present. Need to discuss with NLS as to the likelihood of being able to host such a service by end of the project?		Report for SCURL		**ACTION JT:** Assuming this is also dependent on staffing. Also needs PD
A report of follow on actions for WP2 activity, post project completion	Documented report from WP2 wrap up meeting	**Dec 10**	Report comparing the success of WP2 against intial objectives and proposing follow on/next stage actions.			Blog post/formal project report	
WP3 – Technological Enhancements for Improved Research-centric Functionality and Technical Synergy with the Institutions							

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
User scenarios and use cases for each development module identified (virtual research pool repositories, object description via OAI_ORE, versioning, statistics, HILT integration)	Initial development work now underway at CDLR, still some issues with recruitment at NLS/UoE	End Feb 10	Use cases reviewed by project team for WP3	Internal peer review	Use cases – defined as required following on from WP1 reporting and as required		Completed for some scenarios, however dependent on outcomes of WP1. Work underway with Glasgow University to define use cases for Meprints amendments
Use case realisations/feasibility for each module in Dspace and/or Eprints	Technical designs taken from use cases	Ongoing	Used by development teams, based on scope of use cases identified	Internal peer review	Documented evidence available on project wiki		
Implemented (developed and tested) enhancements for Dspace/Eprints as per requirements	Development of required code. Also 'administrative' technical work required to ensure everyone who is participating is able to participate.(i.e getting final repositories set up etc)						HILT integration into Dspace/Central aggregation and Eprints integration work (based on CDLR activity)
~~Validation report against DCC methodology for designing and evaluating curation and preservation experiments~~	-	-	-	-	-	-	-
~~Validation report against user identified requirements in line with WP1~~	~~Report to review whether or not the initial development proposals meet the actual user~~	~~Required to coincide with development work on project – now~~	~~Honest review of requirements from users (or lack on requirements from~~	~~Peer review~~	-	~~Blog posting – if we build it will they come…..~~	~~NB: This could essentially be the same product as the~~

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
	~~needs as identified through the survey and participation of users~~	~~late~~	~~users!)~~				~~WP1/WP3 evaluation work. Raise change and close product as not happening~~
A report of follow on actions for WP3 activity, post project completion	Account of WP3 development activity, and a documented record of where the developments were left, i.e. what code was completed, what wasn't etc	March 11	Agreed and approved by delivery team	Delivery team review.			
WP4 – Developing an IRIScotland Policy Framework for Organisational and Financial Sustainability							
Common message explaining the business case for OA to Scottish Institutions	Develop with Ian Simpson and Sheila Cannell - basically a revision to the OATS declaration but taking into account new ideas etc	End Now for end March 2011	Approval by initial IRIScotland council meeting group			Formal document, published in conjunction with governance arrangements	Have completed manifesto document
Arranged set of IRIScotland council meetings to discuss OA across project duration	Agree schedule and scope of council activity		Meetings booked, with broad agreed attendance (and support)		Meetings completed and reports posted.		First event scheduled for end November 2010
~~Recurring Scottish 'Mandate watch' report for website in conjunction with SCURL~~	~~Tagged blog postings relating to OA activities~~	-	-	-	-	~~Blog posts~~	~~Deemed unnecessary, based on existing community~~

Comment [JT1]: Pipe dream perhaps, but it would be good to pursue the establishment of an open access mandate for research in Scotlands collection institutions. JT to discuss with Cate Newton? It fits in with the LERU principles.

Specialist outputs	Purpose	Timing	Quality criteria	QA method(s)	Evidence of compliance	Format	Completion notes
							~~postings (i.e. via ROARMAP)~~
Developed website for promotion of IRIScotland council activities to HEI influencers and funders	Using templates and themes from wordpress??	Dependent on IRIScotland council planning				Website (may be on project site)	Have set up, but needs to initially start off as event site
ERIS established cost model for institutional repository set up and operation	To identify costs of repository operations and use data impact of changes in operations against costs/benefits identified					Report Section of WP4 business plan/business case	Based on SDLC experience – perhaps get CW/MW to contribute?
Survey and results of areas of management concern/focus and costs of IR operation		Dedicated SCURL meeting/IRIScotland council meeting?				Section of WP4 business plan/business case	
Report on current costs of repository operations in Scotland	Linked to established costs product			Presentation to SCURL committee		Section of WP4 business plan/business case	
ERIS Proposed business plan for onward operation	Recommendations to be made to SCURL members for how we can take the ERIS work forward. Recommendations for sustainability etc.…					Published report	
ERIS Proposal of governance arrangements for onward operation							